

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 12:39:07 ; Search time 6537 Seconds
(without alignments)
10964.306 Million cell updates/sec

Title: US-10-027-859-1

Perfect score: 1752

Sequence: 1 ttgttttaacaacatgtt.....aaaaaaaaaaaaaaaaaa 1752

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1752	100.0	1752	10	RU91679 U91679 Rattus norv
2	1177.6	67.2	205178	2	AC132020 Rattus no
3	1177.6	67.2	228187	2	AC127107 Rattus no
4	1177.6	67.2	287955	2	AC095777 Rattus no
5	1129	64.4	1489	10	AY049085 Mus muscu
6	841.4	48.0	4956	10	AY049086 Mus muscu
7	841.4	48.0	188987	2	AC138600 Mus muscu
8	841.4	48.0	196584	2	AC104542 Mus muscu
9	841.4	48.0	205624	2	AC139023 Mus muscu
10	841.4	48.0	227724	2	AF336381 Mus muscu
11	764.4	43.6	1805	9	BC023511 Homo sapi
12	760.8	43.4	1901	9	HSR976 H. sapiens m
13	531.2	30.3	169741	9	AC097468 Homo sapi
14	529.6	30.2	1403	9	HSDNAFEV3 Y08979 H. sapiens F
15	296	16.9	1099	9	HSNA38319 Homo sapi
16	292.8	16.7	747	9	HSNA339729 Homo sapi
17	290.2	16.6	742	9	HSNA339670 Homo sapi
18	289.8	16.5	704	9	HSNA340528 Homo sapi
19	280.8	16.0	653	9	HSNA338339 Homo sapi
20	275.2	15.7	880	9	HSNA331239 Homo sapi
21	271.2	15.5	836	9	HSNA339879 Homo sapi
22	270.6	15.4	621	9	HSNA339981 Homo sapi
23	253	14.4	694	9	HSNA325918 Homo sapi
24	233.4	13.3	1754	6	AR232559 Sequence
25	233.4	13.3	1754	6	AR309790 Sequence
26	233.4	13.3	1754	6	BD137836 Hypoxia-r
27	233.4	13.3	1754	6	AF334162 Rattus no
28	233	13.3	730	9	HSNA339666 Homo sapi
29	230.8	13.2	619	9	HSNA339105 Homo sapi
30	229	13.1	616	9	HSNA322575 Homo sapi
31	218.4	12.5	724	9	HSNA339694 Homo sapi
32	211.4	12.1	1436	5	DRE249590 Danio rer
33	210	12.0	596	9	HSNA333530 Homo sapi
34	208.6	11.9	796	9	HSNA339388 Homo sapi
35	205	11.7	1245	5	AF177538 Danio rer
36	204.2	11.7	172173	10	AP003148 Mus muscu
37	203	11.6	816	10	S66169 Erg-3-immun
38	203	11.6	2131	10	AB073078 Mus muscu
39	203	11.6	2133	10	AB073079 Mus muscu
40	203	11.6	2209	10	AB073080 Mus muscu
41	200.4	11.4	762	9	S72621 EWS...erg {
42	200.4	11.4	1219	6	AX657232 Sequence
43	200.4	11.4	1291	6	AX657230 Sequence
44	200.4	11.4	1372	6	AX657234 Sequence
45	200.4	11.4	1509	6	AX657228 Sequence

ALIGNMENTS

RESULT 1
RU91679

LOCUS

DEFINITION

complete cds.

U91679

U91679.1 GI:3033418

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 1752)
Fyodorov,D., Nelson,T. and Deneris,E.

RU91679 1752 bp mRNA linear ROD 07-APR-1998
Rattus norvegicus ETS domain transcription factor Pet-1 mRNA,
complete cds.

TITLE Pet-1, a novel ETS domain factor that can activate neuronal nachr
gene transcription
JOURNAL J. Neurobiol. 34 (2), 151-163 (1998)
MEDLINE 98127904
PUBMED 9468386
REFERENCE 2 (bases 1 to 1752)
AUTHORS Eydorov, D., Nelson, T. and Deneris, E.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-1997) Neurosciences, Case Western Reserve
University, 2109 Adelbert Rd., Cleveland, OH 44106, USA
FEATURES
Location/Qualifiers
1..1752
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/cell_line="PC12"
/tissue_type="pheochromocytoma"
112..1134
/codon_start=1
/product="ETS domain transcription factor PET-1"
/protein_id="AAC12859.1"
/db_xref="GI:3033419"
/translation="MEDPFGAPLGERVPAPHPPOPHPLTAHSSSTPAPGWMQLODP
LPHTLAAASRQALPDPAASLTLPCHQSPRAGITPSAKLTCPVRSPPSPTAQSPA
AMRQSGTQSLILNNYLPDPVGDGLFKRSGSPSMKPLSPAVOKSGGOIQLWQFLLEL
ADRNAGCIIAWEHGHEFKLTDPEVARWGERKSKPMNVYDKLSRLRYVYDKNLS
KVHGKRYAREDFQGLAQACOPPAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA
GLSKLNLMAASAGVAPAGFSYWFGPNATAAAATAALYPTFGUQPPFGVAHAASH
LGHHYH"

polyA_signal 1705..1710 412 g 371 t
BASE COUNT 356 a 613 c 412 g 371 t
ORIGIN

Query Match 100.0%; Score 1752; DB 10; Length 1752;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTTGTTTAAACAAATGTTTATTAGAAAAGTAAATAATTGCATAGTCTTAGTACTTG 60
Db 1 TTTGTTTAAACAAATGTTTATTAGAAAAGTAAATAATTGCATAGTCTTAGTACTTG 60
Qy 61 AACATCAAGTGATTTCAATGAACCGTGAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 120
Db 61 AACATCAAGTGATTTCAATGAACCGTGAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 120
Qy 121 CCAGTGGCGTCTCTGGGGAGAGGTTCCAGCCCCCACCCTCCAGCCCTCCAGCCCTCC 180
Db 121 CCAGTGGCGTCTCTGGGGAGAGGTTCCAGCCCCCACCCTCCAGCCCTCCAGCCCTCC 180
Qy 181 CTCACAGTCACTCTCCAGTACACCGGACCGGATGGGCTGGGATGCAGTCCAGGAC 240
Db 181 CTCACAGTCACTCTCCAGTACACCGGACCGGATGGGCTGGGATGCAGTCCAGGAC 240
Qy 241 CCCCTCCCTCTCAACACACCTGGTCCCGCTCCCGCCAGGCTTTGCCGACCCCGCG 300
Db 241 CCCCTCCCTCTCAACACACCTGGTCCCGCTCCCGCCAGGCTTTGCCGACCCCGCG 300
Qy 301 GCGTCTACTCTTCCCTGTACCCACAGTCACCGGGGGGTATCGGACCCCAAGGCA 360
Db 301 GCGTCTACTCTTCCCTGTACCCACAGTCACCGGGGGGTATCGGACCCCAAGGCA 360
Qy 361 AAGCTGAGTGGCCCCCGGCTGGGCTCCCGCCCATCTCCACCGCCAGTCCCGGAGCG 420
Db 361 AAGCTGAGTGGCCCCCGGCTGGGCTCCCGCCCATCTCCACCGCCAGTCCCGGAGCG 420
Qy 421 ATGAGACAGCGGACCTCCAGCCCTGGTGTATCAACATGTACCTACCATGATCCCGTC 480
Db 421 ATGAGACAGCGGACCTCCAGCCCTGGTGTATCAACATGTACCTACCATGATCCCGTC 480
Qy 481 GGAGATGCTTTTAAAGGAAGGAAGAGCCGAGCTGGGGCGGCTGAGCCCTGCGGTA 540
Db 481 GGAGATGCTTTTAAAGGAAGGAAGAGCCGAGCTGGGGCGGCTGAGCCCTGCGGTA 540

Qy 541 CAGAAAGGACGCGGAGATCCAGTTGTGGCAGTTTCTACTGGAGTGTCTGGCAGACGCG 600
Db 541 CAGAAAGGACGCGGAGATCCAGTTGTGGCAGTTTCTACTGGAGTGTCTGGCAGACGCG 600
Qy 601 GCGAAAGGACGCGGAGATCCAGTTGTGGCAGTTTCTACTGGAGTGTCTGGCAGACGCG 660
Db 601 GCGAAAGGACGCGGAGATCCAGTTGTGGCAGTTTCTACTGGAGTGTCTGGCAGACGCG 660
Qy 661 GACGAGTGGCGGACGCTGGGGCGAGCGAAGAGAGCCCAATATATGAACTACGACAAG 720
Db 661 GACGAGTGGCGGACGCTGGGGCGAGCGAAGAGAGCCCAATATATGAACTACGACAAG 720
Qy 721 CTAAGTGCAGACTGCGCTACTACTACGACAAAACATCATGACGAAAGTGCACGGCAAG 780
Db 721 CTAAGTGCAGACTGCGCTACTACTACGACAAAACATCATGACGAAAGTGCACGGCAAG 780
Qy 781 CGCTACGCTTACCGCTTTGACTTCCAGGGCTGGCAGAGGTTGCCAGGCCACACCCGCG 840
Db 781 CGCTACGCTTACCGCTTTGACTTCCAGGGCTGGCAGAGGTTGCCAGGCCACACCCGCG 840
Qy 841 CAGCCCAACGCGCGCTGCGCGCGCAGCGCAGCGCGCGCCGCGCAGGATGGCGCACTT 900
Db 841 CAGCCCAACGCGCGCTGCGCGCGCAGCGCAGCGCGCGCGCCGCGCAGGATGGCGCACTT 900
Qy 901 TACAAGCTCCCGCTGGTCTGGCTCCACTGCGCTTCCCGCGGCTTCCCAAACTCAACCTT 960
Db 901 TACAAGCTCCCGCTGGTCTGGCTCCACTGCGCTTCCCGCGGCTTCCCAAACTCAACCTT 960
Qy 961 ATGCGAGCTCGGCGCGGCTGGCGCGCGCTGGCTTCTTACTGGCTGGTCCCAACGCC 1020
Db 961 ATGCGAGCTCGGCGCGGCTGGCGCGCGCTGGCTTCTTACTGGCTGGTCCCAACGCC 1020
Qy 1021 ACCGCGCTCGGCGCGCAGCGTGGCTTACCCAAACCGCGGCTTGCAGCCCTCC 1080
Db 1021 ACCGCGCTCGGCGCGCAGCGTGGCTTACCCAAACCGCGGCTTGCAGCCCTCC 1080
Qy 1081 GGGCGCTTTGGCGGCTGGCGCGCTTCCGACTTGGGGGGTCAATATCATAGACGGA 1140
Db 1081 GGGCGCTTTGGCGGCTGGCGCGCTTCCGACTTGGGGGGTCAATATCATAGACGGA 1140
Qy 1141 CGGCGGCTCGAGTGGGCGCTTCCACACAGCGCAGTGACCAATCCCATCTCATCTGG 1200
Db 1141 CGGCGGCTCGAGTGGGCGCTTCCACACAGCGCAGTGACCAATCCCATCTCATCTGG 1200
Qy 1201 GAGAGCGCGAGATTTCCCGGAGCTTCTTACCACAGATTTCTTGCAGCAGCGCT 1260
Db 1201 GAGAGCGCGAGATTTCCCGGAGCTTCTTACCACAGATTTCTTGCAGCAGCGCT 1260
Qy 1261 CCAGCGCCAGGAGAGGATGGGAGCGCTCTGAGGCTTCTTGAATACGAGGCTTCC 1320
Db 1261 CCAGCGCCAGGAGAGGATGGGAGCGCTCTGAGGCTTCTTGAATACGAGGCTTCC 1320
Qy 1321 AGGCTCCCATATATCATCACCCAGGAGGATGATGTCTCCCACTTTAAATTTTCTCTT 1380
Db 1321 AGGCTCCCATATATCATCACCCAGGAGGATGATGTCTCCCACTTTAAATTTTCTCTT 1380
Qy 1381 CCAAGTCTCAGATTTGGAATCTCCCGTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1440
Db 1381 CCAAGTCTCAGATTTGGAATCTCCCGTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1440
Qy 1441 CTTCTCTTTATGACCCCTAGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTT 1500
Db 1441 CTTCTCTTTATGACCCCTAGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTT 1500
Qy 1501 ATTTTCTCTCTCCACGACTACTCCAAACGGTAGTACTCGGTAGTACTCTCGAGGCT 1560
Db 1501 ATTTTCTCTCTCCACGACTACTCCAAACGGTAGTACTCGGTAGTACTCTCGAGGCT 1560
Qy 1561 TCTCACACTCCCTTTTGGGATATGAGAGCATCAAAAACATCTCTGCTGTTGTCATC 1620
Db 1561 TCTCACACTCCCTTTTGGGATATGAGAGCATCAAAAACATCTCTGCTGTTGTCATC 1620
Qy 1621 CTTATCCCAACACTCTGGCTTCCCTTCCATACCACTCTGGGCCCAAGGACCTCTG 1680

```

Db      1621  CCTATCCCAACACTCTGGCTTCGCTCCCTTCCATACCACTCTGCGCCCAAGGACCTCG 1680
|||||
QY      1681  TCTGTATATATCTCTTCAGCCCCCATTAAGATCCCAAGTTCACAAAACAAAAA 1740
|||||
Db      1681  TCTGTATATATCTCTTCAGCCCCCATTAAGATCCCAAGTTCACAAAACAAAAA 1740
|||||
QY      1741  AAAAAAAAAAAAAA 1752
|||||
Db      1741  AAAAAAAAAAAAAA 1752
|||||

RESULT 2
AC132020 205178 bp DNA linear HTG 15-NOV-2002
LOCUS Rattus norvegicus clone CH230-250012, WORKING DRAFT SEQUENCE.
DEFINITION AC132020.3 GI:25007399
ACCESSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
VERSION Rattus norvegicus (Norway rat)
KEYWORDS Rattus norvegicus
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 205178)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barbet,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,R., Hamil,C., Hamilton,C., Hamilton,J.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hollands,B., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,R., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,D., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemelehen,O., Okwuonu,G., Olarnpusagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.,
Puaro,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soes,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Velas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczkyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

```

```

TITLE JOURNAL
REFERENCE JOURNAL
AUTHORS JOURNAL
TITLE JOURNAL
REFERENCE JOURNAL
AUTHORS JOURNAL
TITLE JOURNAL
COMMENT
FEATURES
source
BASE COUNT 51577 a 49768 c 48276 g 50223 t 5334 others
ORIGIN
Query Match 67.2%; Score 1177.6; DB 2; Length 205178;
Best Local Similarity 99.7%; Pred. No. 5e-227;
Matches 1180; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 539 TACCAAGAGCGGCGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTCTGGCAGACC 598
Db 48558 TCCCAACAGGCGGCGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTCTGGCAGACC 48617

```

```

Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 205178)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (29-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 205178)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23196031.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBPD
Center clone name: CH230-250012
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 197555 bases at least Q40
Consensus quality: 198713 bases at least Q30
Consensus quality: 199469 bases at least Q20
Estimated insert size: 203019; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 205178: contig of 205178 bp in length.
FEATURES
1..205178
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-250012"
BASE COUNT 51577 a 49768 c 48276 g 50223 t 5334 others
ORIGIN

```

QY 599 GCGGAAACGCGGCTGATCGCGTGGAGGGCGCGCAGGCTTCAAGCTCAGCGACC 658
Db 48618 GCGGAAACGCGGCTGATCGCGTGGAGGGCGCGCAGGCTTCAAGCTCAGCGACC 48677

QY 659 CCGACGAGTGGCGGCGAGCTGGGCGAGCGCAAGAGCAAGCCCAATATGAATACGACA 718
Db 48678 CCGACGAGTGGCGGCGAGCTGGGCGAGCGCAAGAGCAAGCCCAATATGAATACGACA 48737

QY 719 AGCTAAGTGGAGCACTGGCTACTACTAGCAAAAAATCATGAGCAAGGTGCAAGCA 778
Db 48738 AGCTAAGTGGAGCACTGGCTACTACTAGCAAAAAATCATGAGCAAGGTGCAAGCA 48797

QY 779 AGCGCTACGCTTACCGCTTGAATCTCAGGGGCTGCGACAGGCTTGCAGCACCAACCG 838
Db 48798 AGCGCTACGCTTACCGCTTGAATCTCAGGGGCTGCGACAGGCTTGCAGCACCAACCG 48857

QY 839 CGACGCCACCGCCGCTGCGCGCGCGCGAGCGCGCGCCGCCAGGATGCGCGAC 898
Db 48858 CGACGCCACCGCCGCTGCGCGCGCGCGAGCGCGCGCCGCCAGGATGCGCGAC 48917

QY 899 TTACAAAGTCCCGGCTGGCTTGGCTCCACTGCGCTTCCCGGCTTCCCAAACTCAACC 958
Db 48918 TTACAAAGTCCCGGCTGGCTTGGCTCCACTGCGCTTCCCGGCTTCCCAAACTCAACC 48977

QY 959 TTATGGCAGCCTCGCGCGCGCGCGCGCTGCTTCTTACTGGCTGGTCCCAACG 1018
Db 48978 TTATGGCAGCCTCGCGCGCGCGCGCGCTGCTTCTTACTGGCTGGTCCCAACG 49037

QY 1019 CCACCGCGCTGCGCGCGCGCGCTGCTTACCCAAACCGCGGCTTGCGAGCCCGCTC 1078
Db 49038 CCACCGCGCTGCGCGCGCGCGCTGCTTACCCAAACCGCGGCTTGCGAGCCCGCTC 49097

QY 1079 CCGGCGCTTGGCGCGGCTGCGCGCGCTTCCACTTGGGGGTCTTATCACTAGACGG 1138
Db 49098 CCGGCGCTTGGCGCGGCTGCGCGCGCTTCCACTTGGGGGTCTTATCACTAGACGG 49157

QY 1139 GACGCGCGGTGCACTGGGGCTTCCACACAGCAGCGAGTGCACCAATCCATCTCTCATCT 1198
Db 49158 GACGCGCGGTGCACTGGGGCTTCCACACAGCAGCGAGTGCACCAATCCATCTCTCATCT 49217

QY 1199 GGGAGGAGCCGGAAGATTTCCCGAGTTCCTTTACACAGATTTGCTGACAGCGCG 1258
Db 49218 GGGAGGAGCCGGAAGATTTCCCGAGTTCCTTTACACAGATTTGCTGACAGCGCG 49277

QY 1259 CTCGAGCCAGGGAAGAGATGGGAGGCTCTGAGGTCTCTCTTCAATACGAGGCTT 1318
Db 49278 CTCGAGCCAGGGAAGAGATGGGAGGCTCTGAGGTCTCTCTTCAATACGAGGCTT 49337

QY 1319 CCAGGCTCCCATTTATCATCACCAGGAGGCTGATGCTGCCACATTTAAATTTTCTC 1378
Db 49338 CCAGGCTCCCATTTATCATCACCAGGAGGCTGATGCTGCCACATTTAAATTTTCTC 49397

QY 1379 TTCGAAGTCTCAGATTCGAATCCCGCTTTTTTTTCTCTCTCACTGGAGCCCT 1438
Db 49398 TTCGAAGTCTCAGATTCGAATCCCGCTTTTTTTTCTCTCTCACTGGAGCCCT 49457

QY 1439 GCCTTCTCTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTTCTCTCTCTCTCC 1498
Db 49458 GCCTTCTCTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTTCTCTCTCTCTCC 49517

QY 1499 TCATTTTTTTCTCTCCACGACCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGG 1558
Db 49518 TCATTTTTTTCTCTCCACGACCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGG 49577

QY 1559 CTCTCTACCTCCCTTTTGGGATGAGAGATCAAAAACATCTCTGCTGTGTCCA 1618
Db 49578 CTCTCTACCTCCCTTTTGGGATGAGAGATCAAAAACATCTCTGCTGTGTCCA 49637

QY 1619 TCCCTATCCCAACACTCTGGCTTCCGCTTCCATACCACTCTGCGCCAAAGCCCT 1678
Db 49638 TCCCTATCCCAACACTCTGGCTTCCGCTTCCATACCACTCTGCGCCAAAGCCCT 49697

QY 1679 CGTCTGTATATATCTTTTCAGCCCCCAATTAAGATCCAAGCTC 1722

Db 49698 CGTCTGTATATATCTTTTCAGCCCCCAATTAAGATCCAAGCTTC 49741

RESULT 3
AC127107

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC127107 228187 bp DNA linear HTG 09-NOV-2002
Rattus norvegicus clone CH230-206A13, WORKING DRAFT SEQUENCE.
AC127107
AC127107.3 GI:24818779
HTG: HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 228187)
Murny D, Marie, Metzker, M, Lee, Abranzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, P., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokeme, O., Okwono, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poinexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 228187)
Unpublished
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE
AUTHORS
TITLE
JOURNAL

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 228187)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 9, 2002 this sequence version replaced gi:23264396.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNHC
Center clone name: CH230-206A13
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 216089 bases at least Q40
Consensus quality: 218930 bases at least Q30
Consensus quality: 220275 bases at least Q20
Estimated insert size: 220997; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 'contigs'. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

*
1 228187: contig of 228187 bp in length.
Location/Qualifiers
1..228187
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-206A13"
1..1483
/note="wgs end_extension
clone_end:T7"
1534..2945
/note="wgs end_extension
clone_end:T7"
2996..5407
/note="wgs end_extension
clone_end:T7"
complement(6057..6774)
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH283274"
complement(225072..225776)
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH283275"

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

BASE COUNT 59820 a 53444 c 52556 g 55801 t 6566 others
ORIGIN

Query Match 67.2%; Score 1177.6; DB 2; Length 228187;
Best Local Similarity 99.7%; Pred. No. 5.1e-227;
Matches 1180; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 539 TACAAAGGAGGCGGCGAGATCCAGTTTGTGGCAGTTTCTACTGGAGCTGTGGCAGACC 598
Db 17422 TCCCAACAGGCGGCGAGATCCAGTTTGTGGCAGTTTCTACTGGAGCTGTGGCAGACC 17481

QY 599 GCGGAAAGCGGCGGCGATCCGCTGGAGGGCGGCGGAGTTCAAGCTACCGACC 658
Db 17482 GCGGAAAGCGGCGGCGATCCGCTGGAGGGCGGCGGAGTTCAAGCTACCGACC 17541

QY 659 CCGACGAGGTGGCGGCGAGCTGGGCGGAGCCAAAGCAAGCCCAATATGAATACGACA 718
Db 17542 CCGACGAGGTGGCGGCGAGCTGGGCGGAGCCAAAGCAAGCCCAATATGAATACGACA 17601

QY 719 AGCTAAGTCGAGCACTGCGCTTACTACTACGACAAAAACATCATGAGCAAGGTGCACGCA 778
Db 17602 AGCTAAGTCGAGCACTGCGCTTACTACTACGACAAAAACATCATGAGCAAGGTGCACGCA 17661

QY 779 AGGCTAGCCTACCGCTTTGACTTCCAGGCGCTGGCAGAGCTTTGCCAGCCACCACCG 838
Db 17662 AGGCTAGCCTACCGCTTTGACTTCCAGGCGCTGGCAGAGCTTTGCCAGCCACCACCG 17721

QY 839 CGCAGCCCGCGCGCGCTGCGCGCGGCGGAGGCGGCGCGCCAGGATGCGGCAC 898
Db 17722 CGCAGCCCGCGCGCGCTGCGCGCGGCGGAGGCGGCGCGCCAGGATGCGGCAC 17781

QY 899 TTTCAGAGCTCCCGCGCTGCTGCGCTCCACTGCGCTTCCCGCGGCTCTCCAACTCAACC 958
Db 17782 TTTCAGAGCTCCCGCGCTGCTGCGCTCCACTGCGCTTCCCGCGGCTCTCCAACTCAACC 17841

QY 959 TTATGGAGCTCGGCGCGGCTGCGCGCGGCTTCTTACTGCGCTGGTCCCAAG 1018
Db 17842 TTATGGAGCTCGGCGCGGCTGCGCGCGGCTTCTTACTGCGCTGGTCCCAAG 17901

QY 1019 CCACGCGCGCTGCGCGCGGCGGCGGCTGCGCTTACCAGGCGGCTTGAGCGCCCTC 1078
Db 17902 CCACGCGCGCTGCGCGCGGCGGCGGCTTACCAGGCGGCTTGAGCGCCCTC 17961

QY 1079 CCGGCGCTTTTGGCGCGGCTGCGCGCGGCTTCCGCACTTGGGCGGCTCATTTACTAGAGG 1138
Db 17962 CCGGCGCTTTTGGCGCGGCTGCGCGCGGCTTCCGCACTTGGGCGGCTCATTTACTAGAGG 18021

QY 1139 GACGCGCGGCTGAGTGGGCGCTTCCCAACAGCAGTGAACCAATCCCATCTCTATCT 1198
Db 18022 GACGCGCGGCTGAGTGGGCGCTTCCCAACAGCAGTGAACCAATCCCATCTCTATCT 18081

QY 1199 GGGAGGAGCCCGGAGATTTCCCGGCGGCTTACCAGATTTCCGTTGCAGCAGCG 1258
Db 18082 GGGAGGAGCCCGGAGATTTCCCGGCGGCTTACCAGATTTCCGTTGCAGCAGCG 18141

QY 1259 CTCCAGCCAGGGAAGAGATGGGAAGCTCTGAGGCTTCTCTCAATACAGGCTT 1318
Db 18142 CTCCAGCCAGGGAAGAGATGGGAAGCTCTGAGGCTTCTCTCAATACAGGCTT 18201

QY 1319 CCAGGCTCCCATTTATCATCACTACCCAGGAAGGTCATGTGTCCCACTTTAATTTTCTC 1378
Db 18202 CCAGGCTCCCATTTATCATCACTACCCAGGAAGGTCATGTGTCCCACTTTAATTTTCTC 18261

QY 1379 TTCCAGTCTCCAGATTTCTGGAATCCCGCTTTTCTTCTTCTACCTGGAGCCCT 1438
Db 18262 TTCCAGTCTCCAGATTTCTGGAATCCCGCTTTTCTTCTTCTACCTGGAGCCCT 18321

QY 1439 GCCTTCCTCTTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTCTCTCTCTCTCC 1498
Db 18322 GCCTTCCTCTTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTCTCTCTCTCTCC 18381

QY 1499 TCATTTTTTCTCTCCACGACCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGG 1558
Db 18382 TCATTTTTTCTCTCTCCACGACCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGG 18441

```

Db      18382 TCATTTTCTCTCCAGGACCTACTCAAAACGGTAGTACTCGGTAGTACTCGAGG 18441
OY      1559 CTTCTCACACTCCCTTTTGGGATATGAGAGCATCAAAAACATCTCTGCTGTGTCCA 1618
Db      18442 CTTCTCACACTCCCTTTTGGGATATGAGAGCATCAAAAACATCTCTGCTGTGTCCA 18501
OY      1619 TCCTATCCCAACACTCTGGCTTCGCTCCCTTCATACACACTCTGGCCCAAGGACCT 1678
Db      18502 TCCTATCCCAACACTCTGGCTTCGCTCCCTTCATACACACTCTGGCCCAAGGACCT 18561
OY      1679 CGTCTGTATATATTCCTTTTCAGCCCCCATTAAGATCCAAAGCTTC 1722
Db      18562 CGTCTGTATATATTCCTTTTCAGCCCCCATTAAGATCCAAAGCTTC 18605

RESULT 4
AC095777
LOCUS   Rattus norvegicus clone CH230-9K24, *** SEQUENCING IN PROGRESS ***,
DEFINITION
3 unordered pieces.
AC095777
VERSION AC095777.6 GI:22855924
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE  Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
1 (bases 1 to 287955)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,D., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwono,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sander,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villanana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,

```

```

TITLE JOURNAL
REFERENCE JOURNAL
AUTHORS WILLIAMS,G., WILLSON,R., WLECZYK,R., WOODEN,H., WORLEY,K.,
          WRIGHT,D., WRIGHT,R., WU,J., YAKUB,S., YEN,J., YOON,L., YOON,V.,
          YU,F., ZHANG,J., ZHOU,X., ZHOU,X., ZHAO,S., ZHANG,D., VON
          NIEDERHAUSERN,A., WEISS,R., SMITH,D.R., HOLT,R.A., SMITH,H.O.,
          WEINSTOCK,G. and GIBBS,R.A.
          Direct Submission
          Unpublished
          2 (bases 1 to 287955)
          Worley,K.C.
          Direct Submission
          Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          3 (bases 1 to 287955)
          Rat Genome Sequencing Consortium.
          Direct Submission
          Submitted (03-OCT-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Sep 14, 2002 this sequence version replaced gi:22758717.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
          sequence may extend beyond the ends of the clone and there may be
          contigs that consist entirely of whole genome shotgun sequence
          reads. Both end sequences and whole genome shotgun sequence only
          contigs will be indicated in the feature table.
          ----- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          ----- Project Information
          Center project name: GDBR
          Center clone name: CH230-9K24
          ----- Summary Statistics
          Assembly program: Phrap; version 0.990329
          Consensus quality: 192943 bases at least Q40
          Consensus quality: 197649 bases at least Q30
          Consensus quality: 200917 bases at least Q20
          Estimated insert size: 225772; sum-of-contigs estimation
          Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
          -----
          * NOTE: Estimated insert size may differ from sequence length
          (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          * 1 285136: contig of 285136 bp in length
          * 285137 285236: gap of unknown length
          * 285237 286548: contig of 1412 bp in length
          * 286649 286748: gap of unknown length
          * 286749 287955: contig of 1207 bp in length.
          Location/Qualifiers
            1..287955
              /organism="Rattus norvegicus"
              /mol_type="genomic DNA"
              /db_xref="taxon:10116"
              /clone="CH230-9K24"
            1..1275
              /note="wgs contig"
              /size=27580..29551
              /note="wgs_contig"
          BASE COUNT 51622 a 49444 c 48887 g 53788 t 84214 others
          ORIGIN
          Query Match 67.2%; Score 1177.6; DB 2; Length 287955;
          Best Local Similarity 99.7%; Pred. No. 5.1e-227;

```

Matches 1180; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
QY	539	TACACAAAGGAGCGGGCAGATCCAGTTGTGTCAGTTTCTACTGAGCTGCTGGCAGACC	598
Db	150640	TCCCAACAGGACGCGGAGATCCAGTTGTGTCAGTTTCTACTGAGCTGCTGGCAGACC	150699
QY	599	GCGCAACCGCGGCTGCATCGCTGGGAGGGCGGCACGGGAGTTCAAGCTCACCGACC	658
Db	150700	GCGCAACCGCGGCTGCATCGCTGGGAGGGCGGCACGGGAGTTCAAGCTCACCGACC	150759
QY	659	CGGACGAGTGGCGGAGCTGGGCGAGCGGAGAGCAAGCCCAATATGAACACTAGACA	718
Db	150760	CGGACGAGTGGCGGAGCTGGGCGAGCGGAGAGCAAGCCCAATATGAACACTAGACA	150819
QY	719	AGCTAAGTCGAGCACTGCGCTACTACTACGACAAAACATATGAGCAAGGTGCACGCA	778
Db	150820	AGCTAAGTCGAGCACTGCGCTACTACTACGACAAAACATATGAGCAAGGTGCACGCA	150879
QY	779	AGCGCTACGCTACCGCTTGACTTCCAGGGCTGGCACAGGCTTGCAGCACCAACCG	838
Db	150880	AGCGCTACGCTTACCGCTTGACTTCCAGGGCTGGCACAGGCTTGCAGCACCAACCG	150939
QY	839	CGCAGCCCAACCGCGCTGCGCGCGCGGAGCGGAGCGCCGCCAGATGGCGCAC	898
Db	150940	CGCAGCCCAACCGCGCTGCGCGCGCGGAGCGGAGCGCCGCCAGATGGCGCAC	150999
QY	899	TTTCAAGCTCCCGGCTGGTCTGGCTCCACTGCGCTTCCCGGCGCTTCCAACTCAACC	958
Db	151000	TTTCAAGCTCCCGGCTGGTCTGGCTCCACTGCGCTTCCCGGCGCTTCCAACTCAACC	151059
QY	959	TTATGGCAGCTCGCGCGGCTGGCGCGCGCTGCTTCTTACTGGCTGGTCCCAACG	1018
Db	151060	TTATGGCAGCTCGCGCGGCTGGCGCGCGCTGCTTCTTACTGGCTGGTCCCAACG	151119
QY	1019	CCACCGCGCTGCGCGCGGCTGGCTTACCGAACCGCGGCTGAGCGCCCTC	1078
Db	151120	CCACCGCGCTGCGCGCGGCTGGCTTACCGAACCGCGGCTGAGCGCCCTC	151179
QY	1079	CCGGGCCCTTTGGCGGCTGGCGCGGCTTGGCACTTGGGGGTCATTATCACTAGACGG	1138
Db	151180	CCGGGCCCTTTGGCGGCTGGCGCGGCTTGGCACTTGGGGGTCATTATCACTAGACGG	151239
QY	1139	GACGCGCGGCTGAGTGGGCTTCCACACAGCAGTGCACCAATCCATCCTCATCCT	1198
Db	151240	GACGCGCGGCTGAGTGGGCTTCCACACAGCAGTGCACCAATCCATCCTCATCCT	151299
QY	1199	GGGAGGACCCCGAGATTTCGCCGACGTTCTTTACCACAGATTTCCTTCGACGACGG	1258
Db	151300	GGGAGGACCCCGAGATTTCGCCGACGTTCTTTACCACAGATTTCCTTCGACGACGG	151359
QY	1259	CTCCAGCCAGGGAAGGATGGGAAGCCTCTGAGGCTTCTTCTTGAATACGAGGCTT	1318
Db	151360	CTCCAGCCAGGGAAGGATGGGAAGCCTCTGAGGCTTCTTCTTGAATACGAGGCTT	151419
QY	1319	CCAGGCTCCCATATATCATCACCCAGGAAGGTCATGTCCTCCACATTTATTTTCTC	1378
Db	151420	CCAGGCTCCCATATATCATCACCCAGGAAGGTCATGTCCTCCACATTTATTTTCTC	151479
QY	1379	TTCCAAGTCTCCAGATTCTGGAATCTCCCGTCTTTTCTTCTTCTACCTGGAGCCCT	1438
Db	151480	TTCCAAGTCTCCAGATTCTGGAATCTCCCGTCTTTTCTTCTTCTACCTGGAGCCCT	151539
QY	1439	GCCTTCTCTTTATGACCCCTAGTTTCTGTGTTTTTTTTTTTTTTTTTCTCTCTCTCC	1498
Db	151540	GCCTTCTCTTTATGACCCCTAGTTTCTGTGTTTTTTTTTTTTTTTTTCTCTCTCTCC	151599
QY	1499	TCATTTTTTTTCTCTCCACAGCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGG	1558
Db	151600	TCATTTTTTTTCTCTCCACAGCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGG	151659
QY	1559	CTTCTCACACTCCCTTTTCGGGATATGAGAAGCATCAAAACACTCTCTGCTGTGTCCA	1618
Db	151660	CTTCTCACACTCCCTTTTCGGGATATGAGAAGCATCAAAACACTCTCTGCTGTGTCCA	151719

QY	1619	TCCCTATCCCAACACTCTGGCTTGCCTCCCTTCCATACACACTCTGGCCCAAGGACCT	1678
Db	151720	TCCCTATCCCAACACTCTGGCTTGCCTCCCTTCCATACACACTCTGGCCCAAGGACCT	151779
QY	1679	CGTCTGTATATATTCCTTTTCAGCCCCCATTAAGATCCAAAGTTC	1722
Db	151780	CGTCTGTATATATTCCTTTTCAGCCCCCATTAAGATCCAAAGTTC	151823
RESULT 5			
LOCUS	AY049085	1489 bp	mRNA linear ROD 12-MAR-2002
DEFINITION	Mus musculus ETS-domain transcription factor mRNA, complete cds.		
ACCESSION	AY049085		
VERSION	AY049085.1	GI:19386466	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1489)		
AUTHORS	Pfaar,H., von Holst,A., Vogt Weisenhorn,D.M., Brodski,C., Guimera,J. and Wurst,W.		
TITLE	mPet-1, a mouse ETS-domain transcription factor, is expressed in central serotonergic neurons		
JOURNAL	Dev. Genes Evol.	212 (1), 43-46 (2002)	
MEDLINE	21863846		
PUBMED	11875656		
REFERENCE	2 (bases 1 to 1489)		
AUTHORS	Pfaar,H., von Holst,A., Vogt Weisenhorn,D.M., Brodski,C., Guimera,J. and Wurst,W.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-JUL-2001)	Molecular Neurogenetics, Max-Planck-Institute of Psychiatry, Kraepelinstrasse 2-10, Munich, Bavaria 80804, Germany	
FEATURES	Location/Qualifiers		
source	1..1489		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C3H"		
	/db_xref="taxon:10090"		
	209..922		
	/note="mPet-1"		
	/codon_start=1		
	/product="ETS-domain transcription factor"		
	/protein_id="AAL13055.1"		
	/db_xref="GI:19386467"		
	/translation="MROSGTSQPLLIINMYLPDPVGDGLFKBKSPSWGLSPAVQKGS		
	GOIOLWQFLLELLADRANAGCIAWEGHGFEKLPDPEVARRWGRKSKPNMYDKLS		
	RALRYVDKNTMSKVGKRYAYRFQGLAQACQPPPAHAHAHAHAHAHAHAHAHAHA		
	YKLPAGLAPLPFPGLSKLNLMASAGVAPAGFSYWPGNPATAAATAALYPTPLGP		
	PPGPFPAVAASHLGHYH"		
BASE COUNT	262 a	562 c	378 g 287 t
ORIGIN			
Query Match	64.4%;	Score 1129;	DB 10; Length 1489;
Best Local Similarity	89.5%;	Pred. No. 2.8e-217;	
Matches 1338;	Conservative 0;	Mismatches 120;	Indels 37; Gaps 10;

QY	231	GCTCCAGGACCCCTCCCTCCCTCCACACACCTGGCTGCCCGCTCCCGCAGGCTTGGC	290
Db	18	GCTCCAGGACCCCTCCCTCCCGCCACACCTCGCTGCTCCCGCTTGGC	77
QY	291	GGACCCGCGGCTTACTTCTTCTCCCTGTCAACCCACAGTCACGCGGGGTATCGGCAC	350
Db	78	GGACCCGCGGCTTACTTCTTCTCCCTGTCAACCCACAGTCGCGCGGGATCGGCAC	137
QY	351	CCCAAGCGCAAGCTGACGTG-CCCCCGGTCGGTCCCCCCCATCTCCACCCCGCAGT	409
Db	138	CCCAAGCGCAAGCTGACGTGCCCCCGGTCGATCCCCCCCATCTCCACCCCGCAGT	197
QY	410	CCCCCGGCGGATGAGACAGAGCGGCACCTCCACGCCCCCTGCTGATCAACATGTACTAC	469

3851	ATCGCGGAACGCGCGTTTGCAATCGCGTGGAGGCGGCGCAACGCGAGTTTCAAGCTCACCG	3911
QY	ACCCGACGAGGTGCGCGACGCTGGGCGAGCGCAAGAGCAAGCCCAATATGAACCTACG	715
DB		
3911	ACCCGACGAGGTGCGCGCGCTGGGCGAGCGCAAGAGCAAGCTTAAATGAACTACG	3970
QY	ACAAGCTAAGTCGAGCACTGGCTACTACTACGACAAACACATCATGAGCAAGGTGCACG	775
DB		
3971	ACAAGCTAAGTCGGCGCTGGCTACTACTACGACAAACACATCATGAGCAAGGTGCACG	4030
QY	GCAAGCGCTACGCTACCGCTTTTGACTTTCAGAGGCGCTGGCACAGGCTTGGCAGCCACAC	835
DB		
4031	GCAAGCGCTACGCTACCGCTTTTGACTTTCAGAGGCGCTGGCGAGGCTTGGCAGCCACAC	4090
QY	CCGCGCACCGCCACGCGCGCTGCGCGCGCGAGCGGACGCGCGCGCCAGGATGGCG	895
DB		
4091	CCGCGCAGCCCAACGCGCTGCGCGCGCGCGCGCTGCGCGCGCGCCAGGACGCGC	4150
QY	CACTTTACAGCTCCGGCTGGTCTGGCTCCACTGCGCTTCCCGCGGCTCTCCAAACTCA	955
DB		
4151	CCCTCTACAAGCTCCCGCGCGCTGGCGCGCTGCGCTTCCCGGGCTATCCAAACTCA	4210
QY	ACCTTATGGACCTCCGCGCGCTGGCGCGCTGGCTTCTCTACTGCGCTGGTCCCA	1015
DB		
4211	ACTTGTATGGACCTTCGGCGGGCGTGGCGCGCTGGTTCCTCTACTGCGCTGGCGCGA	4270
QY	ACGCCACCGCGCTCCGCGCGCCACCGCTGCGCTTACCCAAACCCCGGCTTGCAGCGCC	1075
DB		
4271	ACGCCACCGCGCTCCGCGCGCCACCGCGGCTTACCCACCCCGGGCTTGCAGCGCC	4330
QY	CTCCGGGCGCTTTTGGCGGGTGGCGCGCTTCGCACATTGGGGGTCAATTATCACTAGA	1135
DB		
1076	CTCCGGGCGCTTTTGGCGGGTGGCGCGCTTCGCACATTGGGGGTCAATTATCACTAGA	4390
QY	C-GGAGCGCGGCTGCAGTGGGCGCTCTCCACACAG-CAGTGCACCAATCCATCTCTC	1193
DB		
4391	CGGGACGCGCGGTGACAGCGGGGCTCTCTCACATGGCCAGTGACCAATCCCATCTCTC	4450
QY	ATCTTGGGAGGAGCCCGAAGATTCCCGCAGCTTCCCTTTACACACAGATTTCGTTGCAGC	1253
DB		
4451	ATCTTCGAGGAGCCCGAAGCTTTCGCGAGCTTCTTTTAAACAGATTTCATTGCGC	4510
QY	AGCCGCTCCAGCCACGCGAAGAGTGGGAAGCTCTGAGGTCTTCCTTGAATACGA	1313
DB		
4511	AGCCGCGCCACGCGCAGGAAGAAAGATGGGAAACCCCGAGGTCTTCTTGAACATCA	4570
QY	GGCTTCAGGCTCCCATATCATCACCCAGGAAGGTGATGCTGCCCATTTAATTT	1373
DB		
4571	GGCTTCAGGCTCCAGATTCTGGAACCTCCCGT-GTCTTTTCTTTCTCACTTTGCAC	4629
QY	TTCTTTCCAGTCTCCAGATTCTGGAACCTCCCGTCTTTTCTCTCTCACTCTGGAG	1433
DB		
4630	TTCTTTCCAGTCTCCAGATTCTGGAACCTCCCGT-GTCTTTTCTTTCTCACTTTGCAC	4688
QY	CCCTGCTCTCTTTATGACCCCTAGTTTCTGTTTCTTTTCTTTTCTCTCTCTCTC	1493
DB		
4689	CCC---TGACACTCGTTATGACCCCTATTTTCTGGGATTTTCTGCTCTCATTTCTG	4745
QY	TCTCTCATTTTTTTTCTCTCCAGACCTACTCCAAACGGTAGTACCTCGGTAGTACCT	1553
DB		
4746	AC-----TCTCTCCCAACACCCACTCCAAA-----AGGTAGTACCT	4781
QY	CGAGGCTTCTCACACTCCCTTTT-----TCGGGATATGAGAGCATC-AAAAACTCTCTG	1608
DB		
4782	CAAGGCTTCTCGCACTCCCTATTTCTCTCGGATATGAGAGGCTCGATAAACATCTCTG	4841
QY	CTGTGTCCATCCCTATCCCAACACTCTGCTTCTGCTCCCTTCCATACCACACTCTGGCC	1668
DB		
4842	CCGTTCCCATCCCTATATCAACACTCTGCTTCCCTTCTCCCATATCACACTCTGGCC	4901
QY	CAAGGACCTCTGTGTATATATCTTTTCAGGCCCATTAAGATCAA	1717
DB		
4902	CAAGGACCCAGGCTGTATATATCCCTTTTCAGGCCCATTAAGCTCCCA	4950

----- Project Information

Center project name: L28974
Center clone name: 405_L19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 186050 bases at least Q40
Consensus quality: 187399 bases at least Q30
Consensus quality: 187949 bases at least Q20
Insert size: 175000; agarose-fp
Quality coverage: 8.8 in Q20 bases; agarose-fp
Quality coverage: 8.2 in Q20 bases; sum-of-contigs

- * NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 50283: contig of 50283 bp in length
* 50284: gap of 100 bp
* 50384: contig of 1105 bp in length
* 51488: gap of 100 bp
* 51489: contig of 3426 bp in length
* 55014: gap of 100 bp
* 55115: contig of 2641 bp in length
* 57756: gap of 100 bp
* 57855: contig of 7209 bp in length
* 65065: gap of 100 bp
* 65164: contig of 5311 bp in length
* 70475: gap of 100 bp
* 70575: contig of 87947 bp in length
* 158522: gap of 100 bp
* 158623: contig of 20805 bp in length
* 179428: gap of 100 bp
* 179528: 188987: contig of 9460 bp in length.

FEATURES

source

1. 188987
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/map="1"
/clone="RP24-405L19"
/clone_lib="RPCI-24 Male Mouse BAC"

misc_feature

1. 50283
/notes="assembly_fragment"
clone_end:SP6
vector_side:left"

misc_feature

50384..51488
/notes="assembly_fragment"
51589..55014
/notes="assembly_fragment"

misc_feature

55115..57755
/notes="assembly_fragment"
57856..65064
/notes="assembly_fragment"

misc_feature

65165..70475
/notes="assembly_fragment"
70576..158522
/notes="assembly_fragment"

misc_feature

158623..179427
/notes="assembly_fragment"
179528..188987
/notes="assembly_fragment"

misc_feature

clone_end:T7
vector_side:right"
BASE COUNT 50833 a 44352 c 45081 g 47920 t 801 others
ORIGIN

Query Match	48.0%	Score 841.4	DB 2	Length 188987
Best Local Similarity	87.2%	Pred. No. 3.2e-159		
Matches 1037	Conservative 0	Mismatches 116	Indels 36	Gaps 9
Qy	536	CGGTACAGAAAGCAGCGCGGCAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAG	595	
Db	67936	CGTTCCCCACAGCAGCGGCAGATCCAGTTGTGGCAGTTTCTCTGGAGCTGCTGGCAG	67995	
Qy	596	ACCGCGGAACGCGCGCTGTCATCGGTGGAGGGCGGCACGGCGAGTTCAAGCTCACCG	655	
Db	67996	ATCGCGGAACGCGCGTTGTCATCGGTGGAGGGCGGCACGGCGAGTTCAAGCTCACCG	68055	
Qy	656	ACCCGACGAGGTGGCGCGACGCTGGGGCGAGCGCAAGAGACGCCCAATATGAACACG	715	
Db	68056	ACCCGACGAGGTGGCGCGCTGGGGCGAGCGCAAGAGACGCCCAATATGAACACG	68115	
Qy	716	ACAAGCTAGTCAGCACTGCGCTACTACTACACAAAACATCATAGCAAGTGCACG	775	
Db	68116	ACAAGCTAGTCGCGCGCTGCGCTACTACTACACAAAACATCATAGCAAGTGCACG	68175	
Qy	776	GCAAGCGCTACGCTACCGCTTTGACTTCCAGGGCCTGGGCACAGGCTTGCAGCCACAC	835	
Db	68176	GCAAGCGCTACGCTACCGCTTTGACTTCCAGGGCCTGGGCACAGGCTTGCAGCCACAC	68235	
Qy	836	CCGCGCACGCCACGCGCGCTGCGCGCCGCCAGCGGCGCGCGCCAGGATGGCG	895	
Db	68236	CCGCGCACGCCACGCGCGCTGCGCGCCGCCAGCGGCGCGCGCCAGGATGGCG	68295	
Qy	896	CACCTTACAGCTCCCGGCTGGTCTGGCTCCACTGCGCTTCCCGGGCTCTCCAACTCA	955	
Db	68296	CCCTCTACAGCTCCCGGCGGCTTGGCCCGCTGGCCCTTCCCGGGCTTCCAACTCA	68355	
Qy	956	ACCTTATGGCAGCTCGCGCGGCTGGCGCGCCGCTGCTCTTACTGCGCTGGTCCCA	1015	
Db	68356	ACTTGTATGCGAGTTTCGGCGGGGTTTGGCGCCGCTGGTTCTTCTACTGCGCTGGCCGA	68415	
Qy	1016	ACGCCACCGCGCTGCGCGCGCCACCGCTGCGCTTACCCAAACCCCGGGGTTCGAGCCCC	1075	
Db	68416	ACGCCACCGCGCTGCGCGCGCCACCGCGCGCTTACCCAAACCCCGGGGTTCGAGCCCC	68475	
Qy	1076	CTCCCGGGCCCTTGGCGGGTGGCGCGGCTTCGCACCTTGGGGGTTCATATCACTAGA	1135	
Db	68476	CTCCCGGGCCCTTGGCGGGTGGCGCGGCTTCGCACCTTGGGGGTTCATATCACTAGA	68535	
Qy	1136	C-GGGACGGCGGGTGCAGTGGGGCTCTCCACACAG-CGAGTGACCAATCCCATCTC	1193	
Db	68536	CGGGGACGGCGGGTACAGCGGGGCTCTCTCACATGGGCCAGTGACCAATCCCATCTC	68595	
Qy	1194	ATCCTGGGAGGACCCCGAAGATTTCGCCGAGCTTCTTTTACACAGATTTCTGTGGAGC	1253	
Db	68596	ATCCTCGGAGGACCCCGAAGCTTTCGCCGAGCTTCTTTTACACAGATTTCTGTGGAGC	68655	
Qy	1254	AGCGCGTCCACGCGCAGGGAAGAGATGGGAAGCTCTGAGGTCTTCTTCAATACGA	1313	
Db	68656	AGCGCGCGCCACGCGCAGGGAAGAGATGGGAAGCTTCCCGAGGTCTTCTTGAACATCA	68715	
Qy	1314	GGCTTCCAGGCTCCCATTTATCATCACCCAGGAAGGTGCATGCTGCCACTTTAATTT	1373	
Db	68716	GGCTTCCAGGC-CCCGTTATCTTCGCCCCAGGAAGGTGCATGCGGCTCCCATTTAATTT	68774	
Qy	1374	TTCTCTTCCAGTCTCCAGATTTCTGGAACTCCGCTTTTCTTCTTCTTCTTCTTCTTCT	1433	
Db	68775	TTCTCTTCCAGTCTCCAGATTTCTGGAACTCCGCTTTTCTTCTTCTTCTTCTTCTTCT	68833	
Qy	1434	CCCTGCGCTTCTTCTTATGACCCCTAGTTTCTGTTTGTGTTTCTTCTTCTTCTTCTTCT	1493	
Db	68834	CCC---TGCACTCGTGTATGACCCCTATTTTCTGGATTTCTTGGTCTCTCACTTCTG	68890	
Qy	1494	TCTCTCATTTTTTTTCTCTCCACGACCTACTCCAAACGGTAGTACCTCGGTAGTACCT	1553	
Db	68891	AC-----TCTCTCCCAACCCACTCCAAA-----AGGTAGTACCT	68926	

```

QY      1554 CGAGCCTTCTCACATCCCTCTTT-----TCGGGATATGAGAGCATC-AAAACATCTCTG 1608
Db      68927 CAAGGCTTCTCGCACTCCCTATTTCTTCGGGATATGAGAGCCTCGATAAACATCTCTG 68986

QY      1609 CTGTGTGTCATCCCTATCCCAACTCTGGCTTCCTCCCTCCATACCACTCTGCGCC 1668
Db      68987 CCGTTCCTCCATCCCTATATCAACTCTGGCTTCCTCCCTCCATATCACTCTGCGCC 69046

QY      1669 CAAGGACCTCGTCTGTATATATTCCTTTCAGCCCCCATTAAGATCCAA 1717
Db      69047 CAAGGACCCAGCCTGTATATATCTCTTTCAGCCCCCATTAAGATCCCA 69095

RESULT 8
LOCUS   AC104542
DEFINITION AC104542 196584 bp DNA linear HTG 23-MAR-2003
          Mus musculus clone RP23-165D11, WORKING DRAFT SEQUENCE, 9 ordered
          pieces.
ACCESSION AC104542
VERSION   AC104542.3 GI:29164653
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Birren,B., Nusbaum,C. and Lander,E.
          Mus musculus, clone RP23-165D11
          Unpublished
          2 (bases 1 to 196584)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
          Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
          Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
          Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
          Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
          Jones,C., Katat,A., Karatas,A., Kellis,C., LaRocque,K.,
          Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
          Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
          Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
          Direct Submission
          Submitted (13-DEC-2001) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
          3 (bases 1 to 196584)
REFERENCE Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
          Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
          Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
          Collymore,A., Cooke,A., Cooke,P., Corum,B., Dearellano,K.,
          Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fato,S.,
          Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
          Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
          Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Katat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
          Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
          Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
          Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
          Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
          O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
          Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
          Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
          Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,

```

Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viell,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 2003 this sequence version replaced gi:19881898.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: L19847

Center clone name: 165_D_11

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 193071 bases at least Q40

Consensus quality: 194700 bases at least Q30

Consensus quality: 195410 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 195784; sum-of-contigs

Quality coverage: 11.2 in Q20 bases; agarose-fp

Quality coverage: 10.4 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently

* consists of 9 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 21832: contig of 21832 bp in length

* 21833 21932: gap of 100 bp

* 21933 22422: contig of 490 bp in length

* 22423 22522: gap of 100 bp

* 22523 25116: contig of 2594 bp in length

* 25117 25216: gap of 100 bp

* 25217 32792: contig of 7576 bp in length

* 32793 32892: gap of 100 bp

* 32893 38727: contig of 5835 bp in length

* 38728 129564: contig of 90737 bp in length

* 129565 147635: contig of 17971 bp in length

* 147636 147735: gap of 100 bp

* 147736 184494: contig of 36759 bp in length

* 184495 184594: gap of 100 bp

* 184595 196584: contig of 11990 bp in length.

Location/Qualifiers

1..196584

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/clone="RP23-165D11"

/clone_lib="RPC1-23 Female Mouse BAC"

1..21832

/notes="assembly_fragment"

clone_end:SP6

vector_side:left

21933..22422

/notes="assembly_fragment"

22523..25116

/notes="assembly_fragment"

25217..32792

/notes="assembly_fragment"

FEATURES
source

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature 32893..38727
/note="assembly_fragment"
misc_feature 3828..129564
/note="assembly_fragment"
misc_feature 129665..147635
/note="assembly_fragment"
misc_feature 147736..184494
/note="assembly_fragment"
misc_feature 184595..196584
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 51153 a 46239 c 45505 g 52883 t 804 others
ORIGIN
Query Match 48.0%; Score 841.4; DB 2; Length 196584;
Best Local Similarity 87.2%; Pred. No. 3.2e-159;
Matches 1037; Conservative 0; Mismatches 116; Indels 36; Gaps 9;
QY 536 CGGTACAGAAAGGCGGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAG 595
Db |||||
QY 35727 GTTCTCCACAGGAGCGGGAGATCCAGTTGTGGCAGTTTCTCTGGAGCTGTGGCAG 35786
Db |||||
QY 596 ACCGCGGAAAGCGCGGCTGATCGCGTGGAGGGCGGCGAGGAGTTCAAGCTCACCG 655
Db |||||
QY 35787 ATCGCGGAAAGCGCGGTTGATCGCTGGAGGGCGGCGAGGAGTTCAAGCTCACCG 35846
Db |||||
QY 656 ACCCGAGAGGTGGCGGAGCGCTGGGGGAGCGCAAGAGCAAGCCCCAATATGAATACG 715
Db |||||
QY 35847 ACCCGAGAGGTGGCGGCGGCTGGGGGAGCGCAAGAGCAAGCCTTAACATGAATACG 35906
Db |||||
QY 716 ACAAGCTAAGTCAGCACTGCGCTACTACTAGCAGAAACATCATGACGAGGTGCACG 775
Db |||||
QY 35907 ACAAGCTAAGTCGCGCGCTGCGCTACTACTAGCAGAAACATCATGACGAGGTGCACG 35966
Db |||||
QY 776 GCAAGCGTACGCTTACCGCTTTGACTTCCAGGGCTGGCAGAGCTTGCAGCACCAC 835
Db |||||
QY 35967 GCAAGCGTACGCTTACCGCTTTGACTTCCAGGGCTGGCAGAGCTTGCAGCACCAC 36026
Db |||||
QY 836 CGGCGACGCGCCAGCGCGCTGCGCGGCGGAGCGGAGCGCGCCCGCAGGATGCGG 895
Db |||||
QY 36027 CGGCGACGCGCCAGCGCGCTGCGCGGCGGAGCGGAGCTTGCAGCACCACGCGG 36086
Db |||||
QY 896 CACTTTACAAGCTCCCGGCTGGTCTGCTTCCACTGCGCTTCCCGGCGCTTCCAACTCA 955
Db |||||
QY 36087 CCTTCTACAAGCTCCCGGCGGCTTGGCGCGCTTCCCGGCGCTTCCAACTCA 36146
Db |||||
QY 956 ACTTTATGGCAGCTCGCGCGCGTGGCGCGCGCTTCTTCTTACTGGCTGTGTCCTCA 1015
Db |||||
QY 36147 ACTTATGGCAGCTTCGCGCGCGCTTGGCGCGCGCTTCTTCTTACTGGCTGTGTCCTCA 36206
Db |||||
QY 1016 ACGCCACGCGCGCTGCGCGCGCGCTGCGCTTACCCACCGCGGCTTGCAGCGCC 1075
Db |||||
QY 36207 ACGCCACGCGCGCTGCGCGCGCGCTTACCCACCGCGGCTTGCAGCGCC 36266
Db |||||
QY 1076 CTCGCGGCGCGCTTTGGCGCGGTGGCGCGCTTCCACTTGGGGGCTCATTTACTACTAGA 1135
Db |||||
QY 36267 CTCGCGGCGCGCTTGGCGCGGTGGCGCGCTTCCACTTGGGGGCTCATTTACTACTAGA 36326
Db |||||
QY 1136 C-GGACGCGCGGTGAGTGGGCGCTTCCACACAG-CGAGTGAACAAATCCCATCTC 1193
Db |||||
QY 36327 CGGGACGCGCGGTGACAGCGGGGCTCTCATATGGCCAGTGAACAAATCCCATCTC 36386
Db |||||
QY 1194 ATCTGGGAGAGCGCCCAAGATTTCGCCGAGTTCTTTACCAAGATTTCGTTGCGAGC 1253
Db |||||
QY 36387 ATCTGGGAGAGCGCCCAAGATTTCGCCGAGTTCTTTAACACAGATTTCATTTGCGAGC 36446
Db |||||
QY 1254 AGCGCGTCCCGAGCGCGGAGAGGATGGAGCGCTCTAGGCTTCTTCTTGAATACGA 1313
Db |||||
QY 36447 AGCGCGCGCGAGCGCGGAGAGGATGGAGCGCTTCTTCTTGAATACGA 36506
Db |||||
QY 1314 GGCTTCAGGCTCCCATTTATCATCACCCAGAGGGTGCATGTGCTCCCACTTTAATTT 1373
Db |||||

```

```

Db 36507 GGCTTCCAGGC-CCGTTATCTTCGCCCCAGAGAGGTGCATGGCTCCCACTTAATTT 36565
QY 1374 TTCTTTCCAAAGTCTCCAGATTCTGGAACCTCCCGTCTTTTTTTCTTCTCACTGGAG 1433
Db |||||
QY 36566 TTCTTTCCAAAGTCTCCAGATTCTGGAACCTCCCGT-GTCTTTTCTTCTTCTTCTTCTG 36624
Db |||||
QY 1434 CCCCTGCGCTTCTCTTATGACCCCTAGTTTCTGTTTTGTTTTTTTTTTTCTCTCTC 1493
Db |||||
QY 36625 CCC---TGCACCTCGTTATGACCCCTATTTTCTTGGGATTTTCTTGGTCTCATTTCTG 36681
Db |||||
QY 1494 TCTCTCATTTTTTTTCTCTCCACGACTTCTCCAAAGCGGTAGTACCTCGGTAGTACCT 1553
Db |||||
QY 36682 AC-----TCTCTCCACACCCACTCCAA-----AGTAGTACCT 36717
Db |||||
QY 1554 CGAGGCTTCTCACACTCCCTTT----TCGGGATATGAGAAGCATC-AAAAACATCTCTG 1608
Db |||||
QY 36718 CAAGGCTTCTCGCACTCCCTTATTTCTCGGGATATGAGAAGCGCTCGATAAAACATCTCTG 36777
Db |||||
QY 1609 CTGTGTCCATCCCTATCCCAACACTCTGGCTTGGCTTCCCTTCCATACACACTCTGGCC 1668
Db |||||
QY 36778 CCGTTCCCATCCCTATATCAACACTCTGGCTTCCCTTCTCCCATATCACACTCTGGCC 36837
Db |||||
QY 1669 CAAGGACCCCTCTCTGTATATATTTCTTTTTCAGCCCCATTAAGATCCAA 1717
Db |||||
QY 36838 CAAGGACCCCACTGTATATATATCTTTCAGCCCCATTAAGAGCTCCCA 36886
Db |||||

```

```

RESULT 9
AC139023
LOCUS
DEFINITION Mus musculus chromosome 1 clone RP23-430G15 map 1, WORKING DRAFT
SEQUENCE, 6 unordered pieces.
AC139023
AC139023.3 GI:29029286
VERSION HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 1, clone RP23-430G15
Unpublished
2 (bases 1 to 205624)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Meneses,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
REFERENCE
AUTHORS
Submitted (23-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 205624)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Meneses,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

```



```
Db 78151 GCGTTCACAGC-CCGGTATCTTCGCCCCAGGAAGGGTGCATGGGCTCCCACTTTAAATTT 78209
Qy 1374 TTCTCTTCCAAAGTCTCCAGATCTTGGAACTCCCGTCTTTTTTTTCTCTTCTCACCTGGAG 1433
Db 78210 TTCTCTTCCAAAGTCTCCAGATCTTGGAACTCCCGT-GTCTTTTCTTCTTCTCACCTGGAC 78268
Qy 1434 CCCCTGCCCTCTCTTTATGACCCCTAGTTTCTGCTTTTGTGTTTTTTTTTTTCTCTCTCTC 1493
Db 78269 CCC---TGCACTCGTTATGACCCCTATTTTCTGCGGATTTTCTTGGTCTCATTTTCTG 78325
Qy 1494 TCTCTCATTTTTTTTCTCTCCACGACTACTCCAAACGGTAGTACCTCGGTAGTACTCT 1553
Db 78326 AC-----TCTCTCCACACCCACTCCAA-----AGGTAGTACT 78361
Qy 1554 CGAGCTTCTCACACTCCCTTT---TCGGATATGAGAAGCAFC-AAAAACATCTCTG 1608
Db 78362 CAAGGCTTCTCGCACTCCCTTATTTCTCGGATATGAGAAGCTCGATAAACATCTCTG 78421
Qy 1609 CTGTTGTCCATCCCTATCCCACTCTGGCTTCGCTCCCTTCCATACCACTCTGGCC 1668
Db 78422 CGGTTCCCATCCCTATTAACACTCTGGCTTCCTCTTCTCCCATATCACACTCTGGCC 78481
Qy 1669 CAAGGACCTCTGCTGTATATATCTTTCAGCCCCATTAAAGATCCAA 1717
Db 78482 CAAGGACCCAGCCTGTATATATCCCTTCAGCCCCATTAAAGCTCCCA 78530

RESULT 10
AF336381/c 227724 bp DNA linear HTG 02-APR-2001
LOCUS Mus musculus chromosome 1 clone PAC510; PAC457, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AF336381
VERSION AF336381.1 GI:13507298
KEYWORDS HTG; HTGS PHASE1.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 227724)
AUTHORS Rump,A., Hayes,C., Brown,S.D.M. and Rosenthal,A.
JOURNAL Mouse chromosome 1 genomic sequence
REFERENCE 2 (bases 1 to 227724)
AUTHORS Rump,A.
JOURNAL Direct Submission
TITLE Submitted (17-JAN-2001) Genome Analysis, Institute of Molecular
COMMENT Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17869: contig of 17869 bp in length
* 17870 17969: gap of unknown length
* 17970 32746: contig of 14777 bp in length
* 32847 32846: gap of unknown length
* 32847 227724: contig of 194878 bp in length.
FEATURES
Location/Qualifiers
1..227724
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/clones="PAC510; PAC457"
BASE COUNT 57663 a 55423 c 56238 g 58199 t 201 others
ORIGIN
Query Match 48.0%; Score 841.4; DB 2; Length 227724;
```

```
Best Local Similarity 87.2%; Pred. No. 3.2e-159;
Matches 1037; Conservative 0; Mismatches 116; Indels 36; Gaps 9;
Qy 536 CCGTACAGAAAGGAGCGGGCAGATCCAGTTGTGTGACAGTTTCTACTGGAGCTGCTGGCAG 595
Db 144635 CGTTCACAGGAGCGGGCAGATCCAGTTGTGACAGTTTCTCTGGAGCTGCTGGCAG 144576
Qy 596 ACCCGGGAACCGCGGCTGCATCCGCTGGGAGGCGGCACGGCGAGTTCAAGCTCACCG 655
Db 144575 ATCGCGGAACCGCGGCTGCATCCGCTGGGAGGCGGCACGGCGAGTTCAAGCTCACCG 144516
Qy 656 ACCCGAGAGGTGGCGGAGCGCTGGGCGAGCGCAAGAGAACCCCAATATGAACACTAG 715
Db 144515 ACCCGAGAGGTGGCGGCGCTGGGCGAGCGCAAGAGAACCCCAATATGAACACTAG 144456
Qy 716 ACAAGCTAAAGTCGAGCACTGCGCTACTACTACGACAAAAAATCATATGAGCAAGGTGCAC 775
Db 144455 ACNAGCTAAAGTCGCGCGCTGCGCTACTACTAGACAAAAAATCATATGAGCAAGGTGCAC 144396
Qy 776 GCAAGCGCTACGCGCTACCGCTTTCAGCTTCAGGGCTGGGCACAGGCTTGCAGACCACAC 835
Db 144395 GCAAGCGCTACGCGCTACCGCTTTCAGCTTCAGGGCTGGGCAGGCTTGCAGACCACAC 144336
Qy 836 CCGCGACGCGCCACGCGCGCGCTCGCGCGCGGAGGAGCGCGCGCAGAGTGGC 895
Db 144335 CCGCGACGCGCCACGCGCGCTCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 144276
Qy 896 CACTTTAAAGCTCCCGGCTGGTCTGCTGCACCTGCCCTTCCCGCGGCTCTCCAAACTCA 955
Db 144275 CCCTCTAAAGCTCCCGGCGCGCTGCGCGCGCTTCCCGGGCTATCCAAACTCA 144216
Qy 956 ACCTTATGGAGCGCTCGGCGCGGCTGGCGCGCGCTTCTTACTGGGCTGTGTCCA 1015
Db 144215 ACTTGATGGAGCGCTTCGCGCGGCGCTTGGCGCGCGCTTCTTACTGGGCTGTGTCCA 144156
Qy 1016 ACGCCACGCGCGCTCGCGCGCGCGCGCGCTGCGCTCTACCCACCCCGGGCTTGCAGCCC 1075
Db 144155 ACGCCACGCGCGCTCGCGCGCGCGCGCGCTGCGCTCTACCCACCCCGGGCTTGCAGCCC 144096
Qy 1076 CTCCCGGCGCTTGGCGCGGCTGGCGCGCTTCGCACTTGGGGGCTCATTTACACTAGA 1135
Db 144095 CTCCCGGCGCTTGGCGCGGCTGGCGCGCTTCGCACTTGGGGGCTCATTTACACTAGA 144036
Qy 1136 C-GGGACGCGCGGTGAGTGGGCGCTCTCCACACAG-CCAGTGACCAATCCCATCTCTC 1193
Db 144035 CGGGACGCGCGGTACAGCGGGGCTCTCTCATATGGCGCGCTCTCTCATATGGCGCGCTCTC 143976
Qy 1194 ATCTCGGAGGAGCGCCGAGATTTCCCGAGTTTCCCGAGGTTTCTTACACAGATTTCTGTCGAGC 1253
Db 143975 ATCTCGGAGGAGCGCCGAGATTTTCCCGAGGTTTCTTACACAGATTTCTTATTCGAGC 143916
Qy 1254 AGCCGCTCCAGCCAGGGAAGGATGGAAAGCTCTGAGGTCTTCTTGAATACGA 1313
Db 143915 AGCCGCTCCAGCCAGGGAAGGATGGAAAGCTCTTCTTGAATACGA 143856
Qy 1314 GGCTTCAGGCTCCCATATATCATCCCGAGGAGGTGCATGTCTCCCACTTAATTT 1373
Db 143855 GGCTTCAGGCTCCCATATATCATCCCGAGGAGGTGCATGTCTCCCACTTAATTT 143797
Qy 1374 TTCTCTTCCAAAGTCTCCAGATTTCTGGAATCTCCCGCTCTTTTTTTTCTCTCTCTCTCTCTCT 1433
Db 143796 TTCTCTTCCAAAGTCTCCAGATTTCTGGAATCTCCCGT-GTCTTTTTTCTTCTCTCTCTCTCT 143738
Qy 1434 CCCCTGCTCTCTTTATGACCCCTAGTTTCTGTTTTGTTTTTTTTTTTTTCTCTCTCTCTCT 1493
Db 143737 CCC---TGCACCTCGTTATGACCCCTATTTTCTGGGATTTTCTTGGTCTCATTTTCTG 143681
Qy 1494 TCTCTCATTTTTTTTCTCTCCAGGACTACTCCAAACGGTAGTACTCTCGGTAGTACTCT 1553
Db 143680 AC-----TCTCTCCCAACCCACTCCAAA-----AGGTAGTACTCT 143645
Qy 1554 CGAGGCTTCTCACACTCCCTTT---TCGGGATATGAGAAGCATC-AAAAACATCTCTG 1608
```

Db 143644 CAAGGCTTCGCACTCCCTCTATTTCTCGGGATATGAGAGCCCTCGATAAACAATCTCTG 143585
 Qy 1609 CTGTTGTCATCCCTATCCCAACACTCTGGCTTCGGCTCCCTTCATACCACTCTGGCC 1668
 Db 143584 CCGTTCCCATCCCTATATCAACTCTGGCTTCCTTTCTCCCATATCACTCTGGCC 143525
 Qy 1669 CAAGGACCTCGTCTGTATATATTTCTTTAGCCCCCATTAAGATCCAA 1717
 Db 143524 CAAGGACCCAGCCTGTATATATCCCTTTTCAGCCCCCATTAAGATCCCA 143476

RESULT 11

LOCUS

BC023511 1805 bp mRNA linear PRI 23-SEP-2002
 DEFINITION Homo sapiens, FEV protein, clone MGC:12793 IMAGE:4130242, mRNA,
 complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC023511.1 GI:23270720
 MGC.
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1805)

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (05-FEB-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK

COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nhgri.nih.gov
 Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brookes, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,
 Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Teurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL plate: 17 Row: i Column: 5
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 8923788.

FEATURES

source

1..1805
 /location=Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:12793 IMAGE:4130242"
 /tissue_type="Brain, neuroblastoma"
 /clone_lib="NIH MGC 19"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 487..1203
 /codon_start=1
 /product="FEV protein"
 /protein_id="AAH23511.1"
 /db_xref="GI:23270721"
 /db_xref="LocusID:54738"
 /translation="MRQSGASQPLINMVLDPDVGDLFKDGNKPSWGLPSPAVQKGS
 GQIQWQFLLELLADRANACIAWEGHGHEFKLTPDPEVARRWGERKSKPMNYDKLS

CDS

BASE COUNT 250 a 681 c 532 g 302 t
 ORIGIN
 Query Match 43.6%; Score 764.4; DB 9; Length 1805;
 Best Local Similarity 73.1%; Pred. No. 9.5e-144;
 Matches 1110; Conservative 0; Mismatches 381; Indels 27; Gaps 9;
 Qy 240 CCCCCTCCCTCTCACACACACCTTCCCGGCTCCCGCAGGCTTCCCGGACCCGGC 299
 Db 304 CCCCCTCAACCCCGCCAGCCCGGCTTCCCGCAGGCTTCCCGGACCCGGC 363
 Qy 300 GGCGTCTACTTCTCCCTGTACCCACACAGTACACACGGCGGATATCGGCACC-CCAAGCG 358
 Db 364 GCGTCTTCTCTCTCTTGTACCCCGGCTCGCTTCGGGCGGGGATCGGTCCACCGAGCG 423
 Qy 359 CAAAG-CTGACGTGCCCCCGCTCGGTTCCTCCCATCTCCACCGCCAGTCCCGGCA 417
 Db 424 CAAAGCCTGCTCGCCCCCTTCCCGTTCCTCCCATCTCCACCGCCAGTCCCGGCG 483
 Qy 418 GCGATGACAGACGGGACCTCCAGCCCTGCTGTATCAACATGTACTACAGATCC 477
 Db 484 GCGATGACAGACGGGACCTCCAGCCCTGCTGTATCAACATGTACTACAGATCC 543
 Qy 478 GTCCGAGATGCTTTTAAAGGAAGGAAGAGCCGAGCTCGGGGCGCTGAGCCCTCGG 537
 Db 544 GTCGAGACGCTCTCTCAAGGACGGGAAGAACCCGAGCTGGGGGCGCTGAGCCCGCG 603
 Qy 538 GTACAGAAAGCAGCGGGCAGATCCAGTTGTGGCAGTTTCTTACTGGAGCTGTGGCAGAC 597
 Db 604 GTTCAGAAAGCAGCGGACAGATCCAGTGTGGCAGTTTCTTGTGGAGCTGTGGCTGAC 663
 Qy 598 CGCGGAAGCCGCGCTGCATCGCTGGAGGGCGGCACCGCGAGTTCAAGCTTCAACGAC 657
 Db 664 CGCGGAAGCCGCGCTGCATCGCTGGAGGGCGGTCAAGCTTCAAGCTTCAAGCTTCA 723
 Qy 658 CCCGACGAGGTGGCGGCGACGCTGGCGGAGCGCAAGAGCAAGCCCAATATGAATCTACGAC 717
 Db 724 CCGGACGAGGTGGCGGCGGCGGTGGGGAGCGCAAGAGCAAGCCCAATATGAATCTACGAC 783
 Qy 718 AAGCTAAGTGCAGACATCGCTACTACTACGACAAAAACATCATGAGCAAGGTGCACGCG 777
 Db 784 AAGCTAAGTGCAGACATCGCTACTACTACTACGACAAAGCAATCATGAGCAAGGTGCATGCG 843
 Qy 778 AAGGCTTACGCTTACCGCTTTGACTTCAGGGCTTGGCAGAGGCTTGCAGCCACACCC 837
 Db 844 AAGGCTTACGCTTACCGCTTTCGACTTTCAGGGCTTGGCGAGGCTTGCAGCCGCGCC 903
 Qy 838 GCGCACGCCACGCGCGCTGCGCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 897
 Db 904 GCGCACGCTCATGCGCGCGCGCGAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 963
 Qy 898 CTTTACAAGCTCCCGGCTGGTCTGGCTCCACTGCGCTTCCCGCGCTCTCCCAACTCAAC 957
 Db 964 CTCTACAAGTCCCGCGCGGCTCGCGCGCTTCCCGCGCTTCCCGCGCTTCCCAACTCAAC 1023
 Qy 958 CTTATGCGAGCTTGGCGCGGCTGGCGCGCTTCTTCTTACTGGCTTGGTCC--C 1014
 Db 1024 CTATGCGCGCTCGCGCGGCTGCGCGCGCGCGCTTCTCTACTGCGCGCGCGCGCGCG 1083
 Qy 1015 AAGCCACCGCGCTGCGCGCGCGCGCTGCGCTTACCCCAACCCCGGGGCTTGCAGCGCC 1074
 Db 1084 CCGCGCGCGCGCTGCGCGCGCGCGCTTACCCCAAGTCCAGCTTGCAGCGCGCC 1143
 Qy 1075 CTTCCCGGCGCTTGGCGCGGCTGGCGCGCTTTCGACCTTGGGGGCTTATTATCACTAG 1134
 Db 1144 CCGCGCGCGCTTTCGGGCGCGTGGCGCGAGGCTCGGACTTGGGGGCGCATTACCACTAG 1203
 Qy 1135 ACGGACCGCGGCTGAGTGGGGCTTCTCCCAACACAGCCAG-----TGACCAATCCCAT 1189
 Db 1204 ACGGGCGGCTCGGGTGC-CTGCGCGCTCGCGCGCGCGCTAGAGTCTCGCGCGCGATCCCAT 1262

Unpublished (2001)
 3 (bases 1 to 169741)
 Waterston.R.H.
 Direct Submission
 Submitted (18-OCT-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 169741)
 Waterston.R.H.
 Direct Submission
 Submitted (28-NOV-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 169741)
 Waterston.R.H.
 Direct Submission
 Submitted (03-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 6 (bases 1 to 169741)
 Waterston.R.
 Direct Submission
 Submitted (01-MAR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 28, 2001 this sequence version replaced gi:16328295.
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics -----
 Center project name: H_NH0033004
 Drafting Center: WIBR

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:
The RPEC-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatchell, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

SEQUENCING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-64705; the clone sequenced to the right is RP11-559F24. Actual start of this clone is at base position 1 of RP11-330A; actual end is at base position 169741 of RP11-330A.

Data from AC073128, AC092152, and AC020575 was used to finish this clone, AC097468. Polymorphisms have been identified between AC073128, AC020575, and AC097468.

The sequence of AC027000 has been incorporated into AC097468.

FEATURES
Location/Qualifiers
source
1..169741
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-3304"
/clone_lib="RPC1-11"
1..366
/note="match to EST AA237040 (NID:gl861077) zs01g11.s1"
87..270
/note="match to EST BE041205 (NID:g8358326) hk80b10.x1"
495..546
/rpt_family="CT-rich"
567..770
/rpt_family="L1"
788..1089
/rpt_family="Alu"
796..813
/note="match to EST AW206501 (NID:g6505997)"
796..808
/note="match to EST AI697196 (NID:g4985096) tq06h07.x1"
796..808
/note="match to EST AW298786 (NID:g6705422)"
1130..1239
/rpt_family="(TA)n"
1328..1529
/note="match to EST BE041205 (NID:g8358326) hk80b10.x1"
1328..1398
/note="match to EST BI488698 (NID:gl5327926)"
1433..1849
/rpt_family="MaLR"
2100..2463
/rpt_family="MaLR"
2464..2774
/rpt_family="MaLR"
2775..2795
/rpt_family="(T)n"
2782..2797
/note="similar to Homo sapiens EST AW001541 (NID:g5848457) wu34b12.x1"
3079..4360
/rpt_family="MaLR"
4378..4759
/rpt_family="MaLR"
4888..4925
/rpt_family="(ATG)n"
4931..5116
/rpt_family="MIR"
5250..5323
/rpt_family="MIR"
5424..5450
/rpt_family="(A)n"
5490..5562
/rpt_family="L2"
5722..6056
/rpt_family="L2"
6322..6688
/rpt_family="L1"
6696..6851
/rpt_family="MaLR"
6801..6892
/note="match to EST BI488698 (NID:gl5327926)"
7379..7850
/rpt_family="L2"
7910..8326
/rpt_family="MaLR"
8336..8430
/rpt_family="L2"
8464..8578

repeat_region /rpt_family="MIR"
8627..8823
/rpt_family="ERV1"
8830..8917
/rpt_family="L1"
9004..9296
/rpt_family="Alu"
9326..9621
/rpt_family="Alu"
9932..10390
/rpt_family="ERV1"
9996..10275
/note="match to EST BI488698 (NID:gl5327926)"
10393..10460
/rpt_family="(TCTCCC)n"
10403..11260
/rpt_family="SVA"
10561..11681
/note="CpG island (%GC=70.7, o/e=0.70, #CpGs=80)"
11278..12255
/rpt_family="SVA"
12239..12273
/rpt_family="(TAA)n"
12286..12395
/rpt_family="ERV1"
12470..12805
/rpt_family="MaLR"
12752..12781
/rpt_family="(TTTTTG)n"
12808..13349
/rpt_family="L1"
13328..13368
/rpt_family="T-rich"
13350..13660
/rpt_family="Alu"
13350..13368
misc_feature

Query Match 30.3%; Score 531.2; DB 9; Length 169741;
Best Local Similarity 69.3%; Pred. No. 1.2e-96;
Matches 832; Conservative 0; Mismatches 343; Indels 25; Gaps 7;

QY 525 GCTGAGCCTGCGGTACAGAAAGGCGGCGGAGATCCAGTTGTGGCAGTTTCTACTGGA 584
DB 61917 GCTGAGCGCCCTCCCAAGGCGGAGATCCAGTTGTGGCAGTTTCTACTGGA 61858

QY 585 GCTGCTGAGACCGCGCAACCGCGGTGATCGCTGGAGGGCGGCACGGCGAGTT 644
DB 61857 GCTGCTGAGACCGCGCAACCGCGGTGATCGCTGGAGGGCGGCACGGCGAGTT 61798

QY 645 CAAGCTACCGACCCCGGAGGAGTGGCGGAGCGCTGGGCGAGCGCAAGCAAGCCAA 704
DB 61797 CAAGCTACCGACCCCGGAGGAGTGGCGGCGGCTGGGCGAGCGCAAGCAAGCCAA 61738

QY 705 TATGAATACGACAAAGCTAAGTCGAGCACTGCGCTACTACTACGACAAATAATCATGAG 764
DB 61737 CATGAATACGACAAAGCTGAGCGCGCTGCGCTACTACTACGACAAATAATCATGAG 61678

QY 765 CAAGGTGACGGCAAGCGCTACGCTTTCAGTTCAGGCGCTGGGCAAGGCTTG 824
DB 61677 CAAGGTGATGGCAAGCGCTACGCTTTCAGTTCAGGCGCTGGGCAAGGCTTG 61618

QY 825 CCAGCCACCGCGGAGCGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 884
DB 61617 CCAGCCACCGCGGAGCGCGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 61558

QY 885 CCAGGATGGCGCACTTTTAAAGCTCCCGGCTGCTGCTCCAGTCCAGTCCCGCGGCT 944
DB 61557 CCAGGATGGCGCTTACAGGCTGCGCGCGCTGCGCGCTGCGCTCCCGGCT 61498

QY 945 CTCCTAACTCAACTTATGGAGCGCTCGCGCGGCTGGCGCGGCTGCTCTTACTG 1004
DB 61497 CTCCTAACTCAACTTATGGAGCGCTCGCGCGGCTCGCGCGGCTGCTCTTACTG 61438

QY	1005	GCCTGTGTC---CAAAGCCACCGCGCGTGC	CGCCGCCACCGTGTGGCTCTACCCAAACCC	1061
Db	61437	CGCCGGCCCGGGCCCGCCGACCGCTGCGCGCGCA	CGCGCGCGCTCTACCCAGTCC	61378
QY	1062	GGGCTTCAGCCCTCCCGGGCCCTTTGGCGGGT	TGGCGCGCTTCCACTTGGGGG	1121
Db	61377	CAGCTTCGACGCCCGCCCGGGCCCTTCGGGGCG	TGCGCGCGCTCCACTTGGGGG	61318
QY	1122	TCATTATCACTAGACGGACCGCGGGTGCAGT	TGGGGCCCTCTCCACACAGCCAGTGA--	1179
Db	61317	CCATTACCACTAGACGGGGCGTTCGGGTGC-	CTGCGCGCTCCGCCGACCCCTAGAGTCT	61259
QY	1180	---CCAATCCATCTCATCTCGGAGGACCCGAA	GAATTTCCCGACGTTCTCTTACC	1236
Db	61258	CGCCGATCCATCGGATCCCGGGAGGCGCGGAG	CGCTCCGTCAACCGTCTCTAAT	61199
QY	1237	ACGATTTCTTCGACGACCGCTCCCGAGCCAGG	AGAAAGGATGGAAACCTCTGAG	1296
Db	61198	CCAGAGTTTACTCCACTGCGGCACTTAGCAGG	GGGAGC---GSAACGGAAGTCCCTCAA	61142
QY	1297	GTCTTCCTTGAATACAGGAGTTCAGGCTCCCA	TATCATACCCAGGAGGGTGCATG	1356
Db	61141	TCCTTGTCTGTAAGTATTTGCTCTGTCTCCAC	CCCGCAGTCCCTTGAGGAGGCGATG	61082
QY	1357	TGTCGCCACTTT-AAATTTTCTCTCCAAAGTCT	CCAGATTTCTGGAATCTCCCTCTTTT	1415
Db	61081	TGCGCCCTCTTTTCACTTTTTTCTTAGTCTCC	AGGTCCCGAGGGATTTGTGGACC	61022
QY	1416	TTCTCTTCTACCTGAGAGCCCTGCCTCTTTAT	GAACCCCTAGTTTCTGTGTTGTT	1475
Db	61021	TCCTTGTCTCCGCCACACTCCAGTGCATTTCC	GGCTGGCTCTCTAGAAGCCCATCAAT	60962
QY	1476	TTTTTTTTTCTCTCTCTCTCATTTTTTTTCTC	CTCCACGACCTACTCCAAACGGT	1535
Db	60961	ATCACTACTCTTTAAGAGTGCCAAATCTTTCC	CACATTTTGC-TCTCCCAAGAACT	60903
QY	1536	AGTACCTCGGTAGTACTCGAGCTTCTCACACT	CCCCCTTTTCGGGATATGAGAAGCATC	1595
Db	60902	GTCTCCCACTCAGCAGTGGAGGCTCTCAGCGT	CTCTCTCTCTG-----GAC	60854
QY	1596	AAAAACATCTCTGTTGTTCATTCCTATCCACA	CACTCTGGCTTCGCTCCCTTCCATA	1655
Db	60853	CTGAGCAGGTTTGTGAAAGCCACCGCTCTCG	GTGACACACGCGCCCTCTCTCTGTC	60794
QY	1656	CCACTCTGGCCCAAGGACCTCGTCTGTATAT	TCTCTTCAGGCCCATTAAGATCC	1715
Db	60793	CCACTCTCAGAGAACTCCGGTGTGTTCTGACC	CTTTTCAGCCCCCATTAAGCTCC	60734
RESULT	14			
HSDNAFEV3				
LOCUS		1403 bp	DNA	linear
DEFINITION				PRI 23-APR-1997
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REMARK				

Qy	1180	---CCAATCCATCCTCATCTCTGGAGGAGGCCCGAAGATTTC	1236	TCCTTTACC
Db	877	CGCCGATCCATCGGATCTCCGGGAGGCGCCGGAGCCTCCGTCAACCGTCTCTCTAAT	936	
Qy	1237	ACAGATTTCGTGTGAGCAGCGCTCCACAGCCACAGGGAAGAAAGATGGGAAGCCTCTGAG	1296	
Db	937	CCAGAGTTTACTCCACCTGCGCACTTTAGCAGGGGACG---GGACCGAAGCTCCCTCAA	993	
Qy	1297	GTCTTCCTTGAATACGAGGCTTCCAGGCTCCCATATCATCACCCACGGAAGGTCGATG	1356	
Db	994	TCCTTGTCTGGTACTAGATTGTCTCTGTCCACCCCGCAGTCCCTCTGAGGAGGGCGATG	1053	
Qy	1357	TGCTCCCACTTT-AAATTTTTCTCTTCCAAGTCTCCAGATTCTGGAATCTCCGCTCTTTTTT	1415	
Db	1054	TGCGCCCTCTTTCAACCTTTTTTCTCTAGTCTCCAGGTCCTCGAGGGGATTTGTGGACC	1113	
Qy	1416	TTCTCTCTCACCTGAGCGCCTGCCTTCTCTTTATGACCCCTAGTTTTCTGTTTGTGT	1475	
Db	1114	TCCTGTCTCCCGACCACTCCAGTGCATTTCGCGCTGGCTCTCTAGAAGCCCAATCAAT	1173	
Qy	1476	TTTTTTTTTCTCTCTCTCTCATTTTTTTTTTCTCTCCACGACCTACTCTCAAACGGT	1535	
Db	1174	ATCACTACTCTTTAAACGAGTGCCAAATCTTTTCCCACTTTTGC-TCTTCCCAGGNACT	1232	
Qy	1536	AGTACTCGGTAGTACTCGAGGCTTCTCATCTCCGCTTTTCGGGATATGAGAAGCATC	1595	
Db	1233	GCCTCCCACTCAGCACGTGGAGCGCTCTACGGTCTCTCTCTCTGG-----GAC	1281	
Qy	1596	AAAAACAATCTCGCTGTGTTCATTCCTCATCTCCCAACACTCTGGCTTCGCTCCCTCCATA	1655	
Db	1282	CTGAGCAGGTTTGGTGAAGCCACCGTCTCCGTGACACAGGCGCCCTTCTCTCTGTCC	1341	
Qy	1656	CCACACTCTGGCCCCAAGGACCTCGTCTGTATATATTTCTTTTCAGGCCCAATTAAGATCC	1715	
Db	1342	CCACACTCCACAGAGAACTCCGGTGTGTTCTCTGACCCCTTCAGGCCCAATTAAGCTCC	1401	

RESULT 15
HSA338319/c
LOCUS
DEFINITION
1099 bp DNA linear PRI 18-JUL-2002
Homo sapiens genomic sequence surrounding NotI site, clone
NB6-699B.
ACCESSION
AJ338319
VERSION
AJ338319.1 GI:15882737
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
Levisky,V.G., Kolchanov,A.I., Protopopov,A.I., Kashuba,V.I.,
Kisselev,L.B., Wasserman,W., Wahlested,C. and Zabarovsky,E.R.
NotI flanking sequences: a tool for gene discovery and verification
of the human genome
TITLE
Nucleic Acids Res. 30 (14), 3163-3170 (2002)
JOURNAL
MEDLINE
22131767
PUBMED
12136098
REFERENCE
2 (bases 1 to 1099)
AUTHORS
Zabarovsky,E.R.
TITLE
Direct Submission
JOURNAL
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden

Query Match	16.9%	Score 296;	DB 9;	Length 1099;
Best Local Similarity	88.9%;	Pred. No. 2.8e-49;		
Matches 320;	Conservative 0;	Mismatches 40;	Indels 0;	Gaps 0;
525	GCTGAGCCCTCGGTACAGAAAGCAGCGGGCAGATCCAGTGTGTGGCAGTTTCTACTGGA	584		
360	GCTGACCGCCCTCTCCACAGCAGCAGACAGATCCAGCTGTGGCAGTTTCTGCTGA	301		
585	GCTGCTGGCAGACCGCGCGAAACGCGGCTGCATCGCTGGAGGGCGGCACCGCGAGTT	644		
300	GCTGCTGGCTGACCGCGCGAAACGCGGCTGCATCGCTGGAGGGCGGTACCGCGAGTT	241		
645	CAAGCTACACGACCCCGACAGAGTGGCGGCACGCTGGGGCGAGCGCAGAGCAAGCCCAA	704		
240	CAAGCTACAGGACCCGACAGAGTGGCGCGACGGTGGGGCGAGCGCAAGAGCAAGCCCAA	181		
705	TATGAACCTACGACAAAGCTAAGTCGAGCAGCTGCGTACTACTAGCAGCAAAACATCATGAG	764		
180	CATGAACTACGACAAAGCTGAGCGCGCCCTTGGCGTACTACTACGACAAACATCATGAG	121		
765	CAAGGTGACAGGCAAGCGCTACGCGTACCGCTTTGACTTCACGGGGCTTGSCACAGGCTTG	824		
120	CAAGGTGATGGCAAGCGCTACGCGTACCGTTCCGACTTCACGGGCTTGCGCAGGCGCTG	61		
825	CCAGCCACCAACCGCGGACGCGCCACGCGCGCGCTGCCGCCCGCGCGCAGCGCAGCGCGCGC	884		
60	CCAGCGCGCGCGCGACAGCTCATGCGCGCGCGCAGAGTGTCTCCGCGCGCGCGCGCGCGC	1		

Search completed: November 26, 2003, 15:22:22
Job time : 6547 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 12:37:51 ; Search time 492 Seconds
(without alignments)
9612.633 Million cell updates/sec

Title: US-10-027-859-1

Perfect score: 1752
Sequence: 1 ttgttttaacaacacgtt.....aaaaaaaaaaaaaaaaaaaa 1752

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1752	100.0	1752	22	AAS13672
2	1752	100.0	1752	24	ABQ25881
3	1752	100.0	1752	24	ABQ25881
4	318.6	18.2	723	24	ABQ252274
5	318.6	18.2	723	24	ABQ252275
6	262.8	15.0	723	24	ABQ252276
7	262.8	15.0	723	24	ABQ252277
8	233.4	13.3	1754	20	AAQ29137

Handwritten signature/initials

C	9	223	12.7	719	24	ABQ25880	Oligonucleotide fo
10	223	12.7	719	24	ABQ25881	Oligonucleotide fo	
11	223	12.7	2532	22	AA158977	Human polynucleoti	
12	200.4	11.4	1219	25	AB268768	Nucleotide sequenc	
13	200.4	11.4	1291	25	AB268767	Nucleotide sequenc	
14	200.4	11.4	1372	25	AB268769	Nucleotide sequenc	
15	200.4	11.4	1389	25	ABX08759	Angiogenesis-assoc	
16	200.4	11.4	1509	25	AB268766	Nucleotide sequenc	
17	200.4	11.4	3166	25	ABX76295	Lung cancer-associ	
18	199.4	11.4	567	22	ABA48124	Human breast cell	
19	199.4	11.4	567	22	ABA66003	Human foetal liver	
20	199.4	11.4	567	22	ABA33090	Probe #11556 for g	
21	199.4	11.4	567	22	AAK14426	Human brain expres	
22	199.4	11.4	567	22	AAK40160	Human bone marrow	
23	199.4	11.4	567	22	AA120932	Probe #10865 for g	
24	199.4	11.4	567	22	AA146176	Probe #14862 used	
25	199.4	11.4	567	22	AA106643	Probe #6634 used t	
26	199.4	11.4	567	23	ABS39742	Human liver single	
27	199.4	11.4	567	24	ABS14215	Human genome-deriv	
28	198.4	11.3	473	22	ABA50193	Human breast cell	
29	198.4	11.3	473	22	ABA68128	Human foetal liver	
30	198.4	11.3	473	22	ABA35152	Probe #13618 for g	
31	198.4	11.3	473	22	AAK16512	Human brain expres	
32	198.4	11.3	473	22	AAK42265	Human bone marrow	
33	198.4	11.3	473	22	AA148339	Probe #17025 used	
34	198.4	11.3	473	22	AA108696	Probe #8687 used t	
35	198.4	11.3	473	23	ABS41873	Human liver single	
36	195	11.1	1890	14	AAQ50662	Human Hum-Flu-1 ge	
37	194	11.1	2938	14	AAQ50644	Human Hum-Flu-1 ge	
38	194	11.1	2954	22	AAH02915	Human shear stress	
39	194	11.1	2957	24	ABV94304	Breast carcinoma r	
40	194	11.1	2957	24	ABK84139	Human cDNA differe	
41	190.8	10.9	1431	24	ABK73253	DNA encoding human	
42	190.6	10.9	899	23	ABL12097	Drosophila melanog	
43	180	10.3	1447	20	AAK26551	DNA encoding chick	
44	180	10.3	1528	20	AAK26552	DNA encoding chick	
45	178	10.2	2025	23	ABL05371	Drosophila melanog	

ALIGNMENTS

RESULT 1
AAS13672
ID AAS13672 standard; cDNA; 1752 BP.
XX
AC AAS13672;
XX
DT 18-DEC-2001 (first entry)
XX
DE cDNA encoding Rattus norvegicus Pet-1 transcription factor.
XX
KW Rat; EST transcription factor; Pet-1; central serotonin 5-HT neuron;
KW serotonergic receptor; human; animal; cell cycle determined disease;
KW cancer; central nervous system disorder; CNS; psychiatric disorder;
KW neurological disorder; ss.
XX
OS Rattus norvegicus.
XX
PH Key Location/Qualifiers
CDS 112..1134
FT /*tag= a
FT /product= "Pet-1 transcription factor"

US6268216-B1.

31-JUL-2001.

26-JUL-1999; 99US-0360779.

27-JUL-1998; 98US-094264P.

(UYCA-) UNIV CASE WESTERN RESERVE.

XX Deneris ES, Fyodorov DV, Hendricks TJ;
XX WPI; 2001-610396/70.
DR P-PSDB; AAU08495.
XX New oligonucleotide of transcription factor specific for central
PT serotonergic neurons, useful in screening methods for identifying and
PT testing agonists and antagonists of serotonergic activity, comprises
PT DNA sequence from Rattus norvegicus -
XX
PS Claim 1; Fig 1; 34pp; English.
XX The present invention relates to the isolation of a novel transcription
CC factor (Pet-1) from Rattus norvegicus. Pet-1 which is a member of the
CC EST transcription factor family, is specific for central serotonin
CC 5-HT neurons. The sequences of the invention are useful in compositions
CC and methods of identifying and testing serotonergic receptor
CC agonists and antagonists. They are also useful for screening Pet-1
CC binding sites in genomic DNA. The invention is useful in the diagnosis
CC and treatment of a subject (humans or other animals) with cell cycle
CC determined diseases (e.g. cancer), and central 5-HT system disorders
CC such as central nervous system (CNS) disorders, neurological disorders,
CC and psychiatric disorders. The present sequence encodes for Rattus
CC norvegicus Pet-1 transcription factor.
XX
SQ Sequence 1752 BP; 356 A; 613 C; 412 G; 371 T; 0 other;

Query Match 100.0%; Score 1752; DB 22; Length 1752;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTGTTTAAACAACATGTTTATTAGAAAGTAAATAATTGCATAGTCTTAGTACTTG 60
DB 1 TTGTGTTTAAACAACATGTTTATTAGAAAGTAAATAATTGCATAGTCTTAGTACTTG 60

QY 61 AACATCAAGTGATTTATGAAACCGTAGTATCTTCATGTAAACAGTCTTAGATGAAGAC 120
DB 61 AACATCAAGTGATTTATGAAACCGTAGTATCTTCATGTAAACAGTCTTAGATGAAGAC 120

QY 121 CCAGTGGGCTCTCTCTGGGGAGAGGTTCCAGCCGCCATCCATCCATCCATCC 180
DB 121 CCAGTGGGCTCTCTCTGGGGAGAGGTTCCAGCCGCCATCCATCCATCCATCC 180

QY 181 CTCACAGTCTACTCTCTCCAGTACACCGCACCGGATGGGCTGGGATGGGCTCCAGGAC 240
DB 181 CTCACAGTCTACTCTCTCCAGTACACCGCACCGGATGGGCTGGGATGGGCTCCAGGAC 240

QY 241 CCCCTCCCTCTCACACACCTTGCTGCCCTCCCGCAGGCTTGCCGAGCCCGGG 300
DB 241 CCCCTCCCTCTCACACACCTTGCTGCCCTCCCGCAGGCTTGCCGAGCCCGGG 300

QY 301 GGGTCTACTCTCTCTGTCACCCACAGTACACAGGGGGGATCGGACCCCAAGCGCA 360
DB 301 GGGTCTACTCTCTCTGTCACCCACAGTACACAGGGGGGATCGGACCCCAAGCGCA 360

QY 361 AAGCTGAGCTGCCCCCGGCTCCCGCCATCTCCACCGCCAGTCCCGCCGAGCG 420
DB 361 AAGCTGAGCTGCCCCCGGCTCCCGCCATCTCCACCGCCAGTCCCGCCGAGCG 420

QY 421 ATGAGACAGAGCGGACCTCCAGCCCTGCTGATCAACATGTACTACAGATCCCGTC 480
DB 421 ATGAGACAGAGCGGACCTCCAGCCCTGCTGATCAACATGTACTACAGATCCCGTC 480

QY 481 GGAGATGCTCTTTTAAAGAGGAGAGCCGAGCTGGGCGCCCTGAGCCCTCGGTA 540
DB 481 GGAGATGCTCTTTTAAAGAGGAGAGCCGAGCTGGGCGCCCTGAGCCCTCGGTA 540

QY 541 CAGAAAGGCGAGGCGAGATCCAGTTGTGGAGTTTCTACTGGAGCTCTGSCAGACCG 600
DB 541 CAGAAAGGCGAGGCGAGATCCAGTTGTGGAGTTTCTACTGGAGCTCTGSCAGACCG 600

QY 601 GCGAACCGCGGTGCATCGCGTGGAGGGCGGCCACCGCGAGTTCGAAGCTCACCGACCCC 660

DB 601 GCGAACCGCGGTGCATCGCGTGGAGGGCGGCCACCGCGAGTTCGAAGCTCACCGACCCC 660
QY 661 GACGAGGTGGCGCGACCGTGGGGGAGCGCAAGAGCAAGCCCAATATGAACCTACGACAA 720
DB 661 GACGAGGTGGCGCGACCGTGGGGGAGCGCAAGAGCAAGCCCAATATGAACCTACGACAA 720
QY 721 CTAGTCTGAGCACTGCGGTACTACTACGACAAAAACATCATGAGCAAGGTGACGGCAAG 780
DB 721 CTAGTCTGAGCACTGCGGTACTACTACGACAAAAACATCATGAGCAAGGTGACGGCAAG 780

QY 781 CGCTACGCTTACCGCTTTGACTTCCAGGGCTGGCAGAGGTTGGCCAGCCACACCGCG 840
DB 781 CGCTACGCTTACCGCTTTGACTTCCAGGGCTGGCAGAGGTTGGCCAGCCACACCGCG 840

QY 841 CACGCCACCGCGCGTCCCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 900
DB 841 CACGCCACCGCGCGTCCCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 900

QY 901 TACAAGCTCCCGGTGCTGCTGCTCCACTGCGCTTCCCGGGCTCTCCAAACTCAACCTT 960
DB 901 TACAAGCTCCCGGTGCTGCTGCTCCACTGCGCTTCCCGGGCTCTCCAAACTCAACCTT 960

QY 961 ATGGCAGCTCGCGCGCGCTGGCGCGCTGCTGCTTCTTACTGCGCTGGTCCCAACGCC 1020
DB 961 ATGGCAGCTCGCGCGCGCTGGCGCGCTGCTGCTTCTTACTGCGCTGGTCCCAACGCC 1020

QY 1021 ACCGCGCTGCGCGCGCGCTGCGCTTACCCAAACCGCGGGTTGCGAGCCCTCTCC 1080
DB 1021 ACCGCGCTGCGCGCGCGCTGCGCTTACCCAAACCGCGGGTTGCGAGCCCTCTCC 1080

QY 1081 GGGCCCTTTGGCGGGTGGCGCGCTTGGCACTTGGGGGTTCATATCATAGACGGA 1140
DB 1081 GGGCCCTTTGGCGGGTGGCGCGCTTGGCACTTGGGGGTTCATATCATAGACGGA 1140

QY 1141 CGGCGGGTGCAGTGGGGCTCTCCACACAGCAGTACCAATCCCATCTCATCTGG 1200
DB 1141 CGGCGGGTGCAGTGGGGCTCTCTCCACACAGCAGTACCAATCCCATCTCATCTGG 1200

QY 1201 GAGGAGCCCCGAAGATTTCCCGAGTTCCTTTTACCACAGATTTCTGTCAGCAGCGCT 1260
DB 1201 GAGGAGCCCCGAAGATTTCCCGAGTTCCTTTTACCACAGATTTCTGTCAGCAGCGCT 1260

QY 1261 CCCAGCCAGGAGAAAGGATGGAGGCTCTGAGGTCTTCTTGAATACGAGGCTCC 1320
DB 1261 CCCAGCCAGGAGAAAGGATGGAGGCTCTGAGGTCTTCTTGAATACGAGGCTCC 1320

QY 1321 AGGCTCCCATTTATCATCAACCCAGAGGGTGCATGTCTCCCATCTTAAATTTCTCTT 1380
DB 1321 AGGCTCCCATTTATCATCAACCCAGAGGGTGCATGTCTCCCATCTTAAATTTCTCTT 1380

QY 1381 CCAAGTCTCCAGATTTCTGAACTCCCGTCTTTTCTTCTTCTTCTTCTTCTCTCTC 1440
DB 1381 CCAAGTCTCCAGATTTCTGAACTCCCGTCTTTTCTTCTTCTTCTTCTCTCTC 1440

QY 1441 CTTCTCTTTATGAGCCCTAGTCTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1500
DB 1441 CTTCTCTTTATGAGCCCTAGTCTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1500

QY 1501 ATTTTCTTCTCTCCAGCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGGCT 1560
DB 1501 ATTTTCTTCTCTCCAGCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGGCT 1560

QY 1561 TCTCACACTCTCCCTTTTGGGATATGAGAACATCAAAAAATCTCTCTCTGTTGTCATC 1620
DB 1561 TCTCACACTCTCCCTTTTGGGATATGAGAACATCAAAAAATCTCTCTCTGTTGTCATC 1620

QY 1621 CCTATCCCAACACTCTGCTTCCCTTCCATACCACTCTGCGCCCAAGGACCTCTG 1680
DB 1621 CCTATCCCAACACTCTGCTTCCCTTCCATACCACTCTGCGCCCAAGGACCTCTG 1680

QY 1681 TCTGTATATATCTCTTTCAGGCCCATTAAGATCAAGCTTCAAAAAAATAAAAAA 1740
DB 1681 TCTGTATATATCTCTTTCAGGCCCATTAAGATCAAGCTTCAAAAAAATAAAAAA 1740

Db 1681 TCTGTATATATCTCTTTCAGCCCCATTAAAGATCCAAGCTTCAAAAAAAAAAAAAAAAAAAAA 1740
 Qy 1741 AAAAAAAAAAAAA 1752
 Db 1741 AAAAAAAAAAAAA 1752
 RESULT 2
 ID ABQ78589
 XX ABQ78589 standard; cDNA; 1752 BP.
 AC ABQ78589;
 XX 25-NOV-2002 (first entry)
 DT
 XX Nucleotide sequence of transcription factor specific for 5-HT neurons.
 DE
 XX Transcription factor; 5-HT neuron; serotonergic neuron; Pet-1;
 KW ETS domain; gene; ss.
 XX
 OS Bacteriophage lambda 73.
 XX
 XX Key Location/Qualifiers
 PH 112..1134
 FT /*tag= a
 FT /product= "Pet-1"
 FT
 XX US2002090647-A1.
 XX
 XX 11-JUL-2002.
 XX 08-MAY-2001; 2001US-0850799.
 XX 27-JUL-1998; 98US-094264P.
 PR 26-JUL-1999; 99US-0360779.
 XX
 XX (UYCA-) UNIV CASE WESTERN RESERVE.
 PA
 XX Deneris ES, Fyodorov DV, Hendricks TJ;
 XX WPI: 2002-655831/70.
 DR P-PSDB; ABB78207.
 XX
 XX A novel composition for detecting interactive peptide in a sample, and
 PT for detecting serotonergic receptor agonists or antagonists, comprises a
 PT purified Pet-1 peptide
 XX
 XX Disclosure; Fig 1; 30pp; English.
 PS
 XX The present sequence encodes a transcription factor specific for 5-HT
 CC (serotonergic) neurons, designated Pet-1. The Pet-1 protein contains
 CC ETS domains. Compositions comprising Pet-1 are useful for detecting an
 CC interactive peptide, and for identifying and testing seronergic receptor
 CC agonists and antagonists.
 XX
 SQ Sequence 1752 BP; 356 A; 613 C; 412 G; 371 T; 0 other;
 Query Match 100.0%; Score 1752; DB 24; Length 1752;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TTTGTTTTTAAACAACATGTTTATTAGAAAAAGTAAAAATATGTCATAGGTTCTAGTACTTG 60
 Db 1 TTTGTTTTTAAACAACATGTTTATTAGAAAAAGTAAAAATATGTCATAGGTTCTAGTACTTG 60
 Qy 61 AACATCAAGTGATTTTCATGAACCGTGAGTATCTTCATGTAACAGTTCTAGATGGAAGAC 120
 Db 61 AACATCAAGTGATTTTCATGAACCGTGAGTATCTTCATGTAACAGTTCTAGATGGAAGAC 120
 Qy 121 CCAGGTGGCGCTCTCTCTGGGGGAGAGGGTTCCAGCCCCCCCCCTCAGCCCCATCCC 180
 Db 121 CCAGGTGGCGCTCTCTCTGGGGGAGAGGGTTCCAGCCCCCCCCCTCAGCCCCATCCC 180

Qy 181 CTCACAGTCTCACTCTCTCCAGTACACCGGCACCGGGATGGGCTGGGATGCAGCTCCAGGAC 240
 Db 181 CTCACAGTCTCACTCTCTCCAGTACACCGGCACCGGGATGGGCTGGGATGCAGCTCCAGGAC 240
 Qy 241 CCCCTCCCTCTCTACACACACCTGGCTGCCCGCTCCCGCCAGAGCTTGCCTGGACCCCGGG 300
 Db 241 CCCCTCCCTCTCTACACACACCTGGCTGCCCGCTCCCGCCAGAGCTTGCCTGGACCCCGGG 300
 Qy 301 GCGTCTACTCTTCTCCCTGTCAACCCACAGTCAACACGGGGGGGTATCGGCACCCCAAGCGCA 360
 Db 301 GCGTCTACTCTTCTCCCTGTCAACCCACAGTCAACACGGGGGGGTATCGGCACCCCAAGCGCA 360
 Qy 361 AAGCTGACGTGCCCTCCCTCCCTGGCTGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 420
 Db 361 AAGCTGACGTGCCCTCCCTCCCTGGCTGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 420
 Qy 421 ATGAGACAGAGCGGCACCTCCAGCCCCCTTGTGTATCAACATGTACCTACCAAGATCCCTC 480
 Db 421 ATGAGACAGAGCGGCACCTCCAGCCCCCTTGTGTATCAACATGTACCTACCAAGATCCCTC 480
 Qy 481 GGAGATGGTCTTTTAAAGGAAGGAGAGCGCCGAGCTGGGGGGCGCTGAGGCCCTGCGGTA 540
 Db 481 GGAGATGGTCTTTTAAAGGAAGGAGAGCGCCGAGCTGGGGGGCGCTGAGGCCCTGCGGTA 540
 Qy 541 CAGAAAGCGAGCGGCAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGC 600
 Db 541 CAGAAAGCGAGCGGCAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGC 600
 Qy 601 GCGAACCGCGCTGCATCGGTGGAGGGGGCCACCGCGAGTTCAAGCTCACCGACCCC 660
 Db 601 GCGAACCGCGCTGCATCGGTGGAGGGGGCCACCGCGAGTTCAAGCTCACCGACCCC 660
 Qy 661 GACGAGGTGGCGGACGCTGGGGGAGCGGAGAGAGCCCAATATGAATAGACAAAG 720
 Db 661 GACGAGGTGGCGGACGCTGGGGGAGCGGAGAGAGCCCAATATGAATAGACAAAG 720
 Qy 721 CTAAGTCGAGCACTGCGCTACTACTACGACAAAAAATCATGAGCAAGGTGCACGGCAAG 780
 Db 721 CTAAGTCGAGCACTGCGCTACTACTACGACAAAAAATCATGAGCAAGGTGCACGGCAAG 780
 Qy 781 CGTTACCGCTACCGCTTTGACTTCCAGGGCTGCGACAGGCTTCCAGCCACCAACCCCGC 840
 Db 781 CGTTACCGCTACCGCTTTGACTTCCAGGGCTGCGACAGGCTTCCAGCCACCAACCCCGC 840
 Qy 841 CAGCCCCACCGCGCTGCGCGCGCGCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCTT 900
 Db 841 CAGCCCCACCGCGCTGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTT 900
 Qy 901 TACAAGCTCCCGCTGGTCTGGGTCCACTGCGCTTCCCGCGGCTTCCAAACTCAACCTT 960
 Db 901 TACAAGCTCCCGCTGGTCTGGGTCCACTGCGCTTCCCGCGGCTTCCAAACTCAACCTT 960
 Qy 961 ATGGAGCTTGGCGCGCTGGCGCGCGCTGGCGCTTCTTACTGGCTGTGTCCTCAACGCC 1020
 Db 961 ATGGAGCTTGGCGCGCTGGCGCGCGCTGGCGCTTCTTACTGGCTGTGTCCTCAACGCC 1020
 Qy 1021 ACCGCGCTCCCGCGCGCGCGCGCTGCGCTTACCCACCGCGGCTTGCAGCGCCCTCC 1080
 Db 1021 ACCGCGCTCCCGCGCGCGCGCGCTGCGCTTACCCACCGCGGCTTGCAGCGCCCTCC 1080
 Qy 1081 GGGCCCTTTGGCGCGGTGGCGCGCTTCCGACCTTGGGGGGGTCAATATCACTAGACGGGA 1140
 Db 1081 GGGCCCTTTGGCGCGGTGGCGCGCTTCCGACCTTGGGGGGGTCAATATCACTAGACGGGA 1140
 Qy 1141 CGCGCGGGTGCAGTGGGGCTCTCCACACAGCGAGTGACCAATCCCATCTCATCTCTGG 1200
 Db 1141 CGCGCGGGTGCAGTGGGGCTCTCCACACAGCGAGTGACCAATCCCATCTCATCTCTGG 1200
 Qy 1201 GAGAGCCCCGAGAGTTTCCCGACGTTCTTTTACACAGATTTCTGTGAGCAGCGCT 1260
 Db 1201 GAGAGCCCCGAGAGTTTCCCGACGTTCTTTTACACAGATTTCTGTGAGCAGCGCT 1260
 Qy 1261 CCAGCCCGAGGGAAGAAAGATGGGAAGCCCTCTGAGGTCTTCTTGAATAGAGGCTTCC 1320

```

Db 1261 CCAGCCAGGAGAGAGATGGAGGCTCTGAGGCTCTCTTGAATACGAGGCTTC 1320
Qy 1321 AGGCTCCATTATCATCACCCAGGAGGCTGCTGCCACCTTTAATTTTCTCTT 1380
Db 1321 AGGCTCCATTATCATCACCCAGGAGGCTGCTGCCACCTTTAATTTTCTCTT 1380
Qy 1381 CCAAGTCTCAGATCTGGAACTCCCGTCTTTTCTCTCTCTCTCTCTCTCTCTC 1440
Db 1381 CCAAGTCTCAGATCTGGAACTCCCGTCTTTTCTCTCTCTCTCTCTCTCTC 1440
Qy 1441 CTTCCTCTTATGACCCCTAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
Db 1441 CTTCCTCTTATGACCCCTAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
Qy 1501 ATTTTCTCTCTCCAGACCTACTCCAAACGGTAGTACTCGGTAGTACCTCG 1560
Db 1501 ATTTTCTCTCTCCAGACCTACTCCAAACGGTAGTACTCGGTAGTACCTCG 1560
Qy 1561 TCTCAGACTCCCTTTTCGGGATATGAGAGCATCAAAACATCTCTGCTGTTGTC 1620
Db 1561 TCTCAGACTCCCTTTTCGGGATATGAGAGCATCAAAACATCTCTGCTGTTGTC 1620
Qy 1621 CCTATCCCAACTCTGGCTTCGCTCCCTTCCATACCACTCTGGCCCAAGGACCTCG 1680
Db 1621 CCTATCCCAACTCTGGCTTCGCTCCCTTCCATACCACTCTGGCCCAAGGACCTCG 1680
Qy 1681 TCTGTATATTTCTTTCAGCCCATTAAGATCCAAAGCTTCAAAAAA 1740
Db 1681 TCTGTATATTTCTTTCAGCCCATTAAGATCCAAAGCTTCAAAAAA 1740
Qy 1741 AAAAAAAAAA 1752
Db 1741 AAAAAAAAAA 1752

```

RESULT 3

```

AAL38753
ID AAL38753 standard; cDNA; 1752 BP.
XX
AC AAL38753;
XX
DT 02-SEP-2002 (first entry)
XX
DE Rat lambda73 cDNA sequence.
XX
KW Seronergic receptor activity; binding site; serotonergic gene; Pet-1;
KW transcription factor; serotonin; 5-HT; neuron; basal transcription;
KW autonomous enhancer element; rat; lambda73; gene; ss.
XX
OS Rattus norvegicus.

```

```

FH Key Location/Qualifiers
FT CDS 112..1134
FT FT /*tag= a
FT FT /product= "Rat lambda73 protein"
XX

```

US6384204-B1.

07-MAY-2002.

05-NOV-1999; 99US-0435335.

27-JUL-1998; 98US-094264P.

26-JUL-1999; 99US-0360779.

(DENE/) DENERIS E S.

(FYOD/) FYODOROV D V.

(HEND/) HENDRICKS T J.

Deneris ES, Fyodorov DV, Hendricks TJ;

WPI; 2002-488197/52.

```

DR P-PDB; AAO21529.
XX
PT New oligonucleotide that binds to transcription factor Pet-1, specific
PT for serotonergic neurons, functions as autonomous enhancer element -
PS Claim 2; Fig 1; 35pp; English.
XX
CC The invention relates to isolated, purified oligonucleotides having seven
CC specified sequences defined in the specification. The oligonucleotides of
CC the invention can be used for screening compounds that are agonistic or
CC antagonistic to seronergic receptor activity. The oligonucleotides are
CC binding sites to seronergic genes for the new transcription factor
CC Pet-1, specific for serotonin (5-HT) neurons. Binding of Pet-1 to the
CC 5-HT neurons supports transcription or stimulates basal transcription,
CC i.e. they function as autonomous enhancer elements. This polynucleotide
CC sequence represents a rat lambda73 cDNA sequence which encodes a
CC transcription factor specific for central serotonin 5-HT neurons of the
CC invention.
XX
SQ Sequence 1752 BP; 356 A; 613 C; 412 G; 371 T; 0 other;

```

```

Query Match 100.0%; Score 1752; DB 24; Length 1752;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 TTTGTTTTAAACAACATGTTTATTAGAAAAGTAAATAATTGTCATAGGCTTTAGTACTTG 60
Db 1 TTTGTTTTAAACAACATGTTTATTAGAAAAGTAAATAATTGTCATAGGCTTTAGTACTTG 60
Qy 61 AACATCAAGTGTTATTCATGAACCGTGAGTATCTTCATGTAACAGTCTTAGATGGAAGAC 120
Db 61 AACATCAAGTGTTATTCATGAACCGTGAGTATCTTCATGTAACAGTCTTAGATGGAAGAC 120
Qy 121 CCAGGTGGCGCTCTCTGGGGGAGAGGGTTCCAGAGCCCCCACCCTCCAGCCCCCATCCC 180
Db 121 CCAGGTGGCGCTCTCTGGGGGAGAGGGTTCCAGAGCCCCCACCCTCCAGCCCCCATCCC 180
Qy 181 CTCAGAGCTCACTCTCTCAGTACACCGGCACCGGATGGCTGGGATGAGTCCAGGAC 240
Db 181 CTCAGAGCTCACTCTCTCAGTACACCGGCACCGGATGGCTGGGATGAGTCCAGGAC 240
Qy 241 CCCCTCCCTCTCCACACACCTGCTGCCGCTCCCGCCAGGCTTCCCGGACCCCGGG 300
Db 241 CCCCTCCCTCTCCACACACCTGCTGCCGCTCCCGCCAGGCTTCCCGGACCCCGGG 300
Qy 301 CGGTCTACTCTTCCCTGTACCCACAGTACACACCGGCGGG'ATCGGCACCCCAAGCGCA 360
Db 301 CGGTCTACTCTTCCCTGTACCCACAGTACACACCGGCGGGTATCGGCACCCCAAGCGCA 360
Qy 361 AAGCTGAGTGGCCCCCGCTGGGTTCCTCCCATCTCCACCGCCAGTCCCGGCGAGCG 420
Db 361 AAGCTGAGTGGCCCCCGCTGGGTTCCTCCCATCTCCACCGCCAGTCCCGGCGAGCG 420
Qy 421 ATGACACAGAGCGGCACCTCCCGAGCCCTGCTGATCAACATGACTACACAGTCCCGTC 480
Db 421 ATGACACAGAGCGGCACCTCCCGAGCCCTGCTGATCAACATGACTACACAGTCCCGTC 480
Qy 481 GGAGATGCTCTTTTAAAGGAAGGAGAGCCGAGCTGGGGGCCGCTGAGCCCTCGGTA 540
Db 481 GGAGATGCTCTTTTAAAGGAAGGAGAGCCGAGCTGGGGGCCGCTGAGCCCTCGGTA 540
Qy 541 CAGAAAGCGAGCGGCAGATCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGC 600
Db 541 CAGAAAGCGAGCGGCAGATCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGC 600
Qy 601 GCGAACGCGCGCTGCTGCTGGAGGGCGGCACCGAGTTCAGTCAAGCTCACCCACCCC 660
Db 601 GCGAACGCGCGCTGCTGCTGGAGGGCGGCACCGAGTTCAGTCAAGCTCACCCACCCC 660
Qy 661 GACGAGGTGGCGCGACGCTGGGGGAGCGAGCGCAAGAGCAAGCCCAATATGAACTACGACAAG 720
Db 661 GACGAGGTGGCGCGACGCTGGGGGAGCGAGCGCAAGAGCAAGCCCAATATGAACTACGACAAG 720

```

QY 721 CTAAGTCGAGCACTGGCTACTACTACGACAAAAACATCATGAGCAAGGTGACGGCAAG 780
Db |||||
QY 721 CTAAGTCGAGCACTGGCTACTACTACGACAAAAACATCATGAGCAAGGTGACGGCAAG 780
Db |||||
QY 781 CGCTACGCTACCGCTTTGACTTCAGGGCTGGACAGGCTTGCAGCCACACCCGG 840
Db |||||
QY 781 CGCTACGCTACCGCTTTGACTTCAGGGCTGGACAGGCTTGCAGCCACACCCGG 840
Db |||||
QY 841 CACGCCACGCGCGCTGCGCCGCGCAGCGGAGCGCGCCAGGATGGCGACTT 900
Db |||||
QY 841 CACGCCACGCGCGCTGCGCCGCGCAGCGGAGCGCGCCAGGATGGCGACTT 900
Db |||||
QY 901 TACAGCTCCCGCTGCTGCTCCACTGCGCTTCCCGGCTCTCCTCAAACTCAACTT 960
Db |||||
QY 901 TACAGCTCCCGCTGCTGCTCCACTGCGCTTCCCGGCTCTCCTCAAACTCAACTT 960
Db |||||
QY 961 ATGGCAGCTCCGCGCGCTGCGCCGCTGCTTCTTACTGCGCTGGTCCCAACGCC 1020
Db |||||
QY 961 ATGGCAGCTCCGCGCGCTGCGCCGCTGCTTCTTACTGCGCTGGTCCCAACGCC 1020
Db |||||
QY 1021 ACCGCGCTGCGCGCCACCGCTGCTGTACCAACCCCGGCTTGACGCCCTCC 1080
Db |||||
QY 1021 ACCGCGCTGCGCGCCACCGCTGCTGTACCAACCCCGGCTTGACGCCCTCC 1080
Db |||||
QY 1081 GGGCCCTTTGGCGGCTGGCGCGCTTGGCACTTGGGGGTCAATATCACTAGACGGA 1140
Db |||||
QY 1081 GGGCCCTTTGGCGGCTGGCGCGCTTGGCACTTGGGGGTCAATATCACTAGACGGA 1140
Db |||||
QY 1141 CGCGCGGCTGAGTGGGGCTCTCCACACAGCAGTGACCAATCCCATCTCTCTCTGG 1200
Db |||||
QY 1141 CGCGCGGCTGAGTGGGGCTCTCCACACAGCAGTGACCAATCCCATCTCTCTGG 1200
Db |||||
QY 1201 GAGGAGCCCGAAGATTTCCCGAGCTTCCACACAGATTTCTGTTGACGACCGCT 1260
Db |||||
QY 1201 GAGGAGCCCGAAGATTTCCCGAGCTTCCACACAGATTTCTGTTGACGACCGCT 1260
Db |||||
QY 1261 CCCAGCCAGGAGGAGGATGGAGGCTCTGAGGCTTCTTGAATACGAGGCTCC 1320
Db |||||
QY 1261 CCCAGCCAGGAGGAGGATGGAGGCTCTGAGGCTTCTTGAATACGAGGCTCC 1320
Db |||||
QY 1321 AGGCTCCCATATATCATATCCCGAGGCTGCTGCTCCACTTTAAATTTCTCTT 1380
Db |||||
QY 1321 AGGCTCCCATATATCATATCCCGAGGCTGCTGCTCCACTTTAAATTTCTCTT 1380
Db |||||
QY 1381 CCAAGTCTCCAGATTCGGAATCCCGCTTTTTTTTCTCTCTCACTGAGCCCTGC 1440
Db |||||
QY 1381 CCAAGTCTCCAGATTCGGAATCCCGCTTTTTTTTCTCTCTCACTGAGCCCTGC 1440
Db |||||
QY 1441 CTTCCTTTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTTTCTCTCTCTCTC 1500
Db |||||
QY 1441 CTTCCTTTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTTTCTCTCTCTCTC 1500
Db |||||
QY 1501 ATTTTCTTTCTCCAGGACCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGCT 1560
Db |||||
QY 1501 ATTTTCTTTCTCCAGGACCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGCT 1560
Db |||||
QY 1561 TCTACACTCCCTTTTGGGATATGAGAGCATCAAAAATCTCTCTGTGTGTCATC 1620
Db |||||
QY 1561 TCTACACTCCCTTTTGGGATATGAGAGCATCAAAAATCTCTCTGTGTGTCATC 1620
Db |||||
QY 1621 CCTATCCCAACACTCTGCTTCCCTCCCTCCATACCACTCTTGCCCAAGGACCTCG 1680
Db |||||
QY 1621 CCTATCCCAACACTCTGCTTCCCTCCCTCCATACCACTCTTGCCCAAGGACCTCG 1680
Db |||||
QY 1681 TCTGTATATTTCTTTTCCAGCCCAATTAAGATCCAAAGCTTCAAAAAA 1740
Db |||||
QY 1681 TCTGTATATTTCTTTTCCAGCCCAATTAAGATCCAAAGCTTCAAAAAA 1740
Db |||||
QY 1741 AAAAAA 1752
Db |||||
QY 1741 AAAAAA 1752
Db |||||

RESULT 4
ABQ52274/c
ID ABQ52274 standard; DNA; 723 BP.
XX
AC ABQ52274;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38865.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 723 BP; 94 A; 89 C; 301 G; 239 T; 0 other;
Query Match 18.2%; Score 318.6; DB 24; Length 723;
Best Local Similarity 70.3%; Pred. No. 9.7e-54;
Matches 442; Conservative 0; Mismatches 184; Indels 3; Gaps 1;
QY 526 CTGAGCCCTCGGTACAGAAAGGCGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAG 585
Db |||||
QY 586 CTGCTGGCAGACCGGCAACCGGCTGCTCGGTGGGAGGGCGGCGGCGGAGTTC 645
Db |||||
QY 647 CTACTAACTAACCGCGGACCGGCTACATCGGTAAAAAACCACGATCACGACGATTC 588
Db |||||

```

QY 646 AAGCTCACGACCCGACGAGGTGGCGGAGCGCTGGGGGAGCGCAAGAGAGCCCAAT 705
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 587 AACTCAGNACCCGACGAGGATACGCGAGGATATAAAGCAAGCAAAACCAAC 528
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 706 ATGAATACGACAAAGTAAAGTCGAGCACTGCGCTACTACTACGACAAAACATCATGAGC 765
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 527 ATAACTACGACAAACTAAACCGCGCTACGCTACTACTACGACAAAACATCATAAAC 468
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 766 AAGGTACGCGGACGCTACGCTACCGCTTTGACTTCCAGGCGCTGGCAGGTTGC 825
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 467 AAAATACATAAACAACGCTACGCTTACCGCTTCGACTTCCAAACCTAACGCAAACTAC 408
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 826 CAGCCACACCCCGGACGCGCCACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 885
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 407 CAACCGCGCGCGCGCGCGCGCTACATACGCGCGCGCGCGCGCGCGCGCGCGCGCG 348
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 886 CAGGATGGCGCACTTACAAAGTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 347 CAAAACGAGCGCTTACAAACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 288
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 946 TCCAAACTCAACCTTATGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1005
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 287 TCCAAACTCAACCTTATGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 228
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1006 CTGTGTCCTCCAA---CGCCACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1062
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 227 CCGAACCGGACCCCGCGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 168
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1063 GGTTCGAGCGCGCTCCCGCGCGCGCTTGGCGCGCGCTTGGCGCGCGCTTGGCGCG 1122
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 167 AACTTACAAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 108
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1123 CATTATCACTAGACGGAGCGCGCGCGGTGC 1151
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 107 CATTACCACTAAACGAAACGATCGAATAC 79
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 5
ABQ52275
ID ABQ52275 standard; DNA; 723 BP.
XX
AC ABQ52275;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38866.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-BP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
WPI: 2002-371829/40.
XX
DR
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX

```

```

PS
XX
Claim 12; 56pp + Sequence Listing; 56pp; German.
This invention describes a novel method for determining the degree of
methylation of a particular cytosine in a motif 5'-CpG-3', present in a
genomic sample of DNA. The sample is treated chemically to convert
cytosine (C) but not methylated C to uracil, then part of the genomic
DNA that contains the target C is amplified to form a labeled amplicon.
The amplicon is hybridised to two classes, each with at least one
member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
and the degree of hybridisation to both classes is determined from the
label on the amplicon. From the ratio of labels hybridised to the two
classes of oligomers, the degree of methylation is calculated. The method
is used: (i) for diagnosis and/or prognosis of side effects of
therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
of the central nervous, cardiovascular, gastrointestinal and respiratory
systems etc., particularly by detecting mutations or single nucleotide
polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
types and for investigating cell differentiation. The method allows the
methylation status of many C residues to be determined simultaneously.
ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the
method for determining the degree of cytosine methylation described in
the disclosure of the invention.
XX
Sequence 723 BP; 239 A; 301 C; 89 G; 94 T; 0 other;
Query Match 18.2%; Score 318.6; DB 24; Length 723;
Best Local Similarity 70.3%; Pred. No. 9.7e-54;
Matches 442; Conservative 0; Mismatches 184; Indels 3; Gaps 1;
QY 526 CTGAGCCCTGCGGTACAGAAAGCGGAGCGGAGTCCAGTTCGTGGCAGTTCTTACTGGAG 585
Db 17 CTAAACGCGCCCTCCCAACAAACGAAACAAATCCAACTATACAAATTTCTACTAAAA 76
QY 586 CTGCTGGCAGACCGCGGAAACCGCGCTGCATCGGTGGGAGCGCGCGCGAGTTTC 645
Db 77 CTACTAACTAACCGCGGAAACCGCGCTACATCGGTAAAAAAGCATCAGCAGAAATTC 136
QY 646 AAGCTCACGACCCGACGAGGTGGCGGAGCGCTGGGGGAGCGCGCAAGAGAGCCCAAT 705
Db 137 AAACCTCAGCAACCCCGAAGCAAAATAACGCGAGATAAACGAAACGAAACCAACCAAC 196
QY 706 ATGAATACGACAAAGTAAAGTCGAGCACTGCGCTACTACTACGACAAAACATCATGAGC 765
Db 197 ATAACTACGACAAACTAAACCGCGCTTACCTACTACTACGACAAAACATCATAAAC 256
QY 766 AAGGTGACGCGCAAGCGCTACGCTTACCGCTTTGACTTCCAGGCGCTGGCAGCGTTGC 825
Db 257 AAAATACATAAACAACGCTACGCTTACCGTTTCGACTTCCAAACCTTACGCAAACTAC 316
QY 826 CAGCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885
Db 317 CAACCGCGCGCGCGCGCGCGCTCATACCGCGCGCGCAACTACTACCGCGCGCGCGCGCG 376
QY 886 CAGGATGGCGCACTTTTACAAGCTCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
Db 377 CAACGAGCGCGCTCTACAAACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 436
QY 946 TCCAAACTCAACCTTATGCGAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1005
Db 437 TCCAAACTCAACCTTATGCGAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496
QY 1006 CTGTGTCCTCCAA---CGCCACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
Db 497 CCGAACCCGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 556
QY 1063 GGTTCGAGCGCGCTCCCGCGCGCGCTTTGGCGCGGTGGCGCGGTTCGCGCTTGGGGGT 1122
Db 557 AACTTACAAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 616
QY 1123 CATTATCACTAGACGGAGCGCGCGGTGC 1151
Db 617 CATTACCACTAAACGAAACGATCGAATAC 645

```

Qy	664	GAGGTGGCGGACGCCTGGGGCGAGCCAAAGACAAGCAAGCCCAATATGAACCTACGACAGCCTA	723
Db	155	GAGTTCGGCGCGCGTGCGGGCGAGCGCTTAAGAGTTAAATTGAATTACGATAAGTTG	214
Qy	724	AGTCGAGCACCTGGCTACTACTACGACCAAACAACATCATGAGCAAGGTGCGACGCCAACGCCG	783
Db	215	AGTCGGGTGTTTTGCGTTATTTATACGATAAGAATATATGAGTAAGGTGTATGTAAGCGT	274
Qy	784	TACGCCCTACC CGCTTTTGACTTTCAGGGCCCTGGCACAGGCTTTGCCAGGCCACCAACCCGCGCAC	843
Db	275	TACGTTTATCGTTTTTCGATTTTTAGGGCTTTGGCGTAGGTTGTTAGTCGTCGTTCCGCGTAC	334
Qy	844	GCCACGCGCGCGCTCGCGCGCGCGCGCAGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTTTAC	903
Db	335	GTTTATATGTCGTCGTCGTA GTTGTGTCGTCGTCGCGGTCGTTTTTAGGACGCGCGGTTTTAT	394
Qy	904	AAGCTCCC GCGCTGCTGGCTCCACTGCCCTCCCCGCGCTCTCCAACCTCAACCTTATG	963
Db	395	AAGTTGTTGTCGCGTTTCGTTTCGTTGTTTTTTTCGTTTTTTAAAATTAATTATG	454
Qy	964	GCAGCTCGCGCGCGTGCGCGCGCTGCTCTCTTACTGGCCTGGTCC--CAACGCC	1020
Db	455	GTCGTTTTCGTCGCGGTCGCTGTCGCGTTTTTTTATTTGTCGCGGTTTCGGGTTTCGTC	514
Qy	1021	ACGCGCTGCGCGCGCGCACCGCTGCGCTCTACCCAAACCCCGGGCTTGCAGCCCCCTCCC	1080
Db	515	GTTATCGTTGTCGTCGCTATATCGTCGCGTTTTATTATTAGTTTTGATGTTTTTCGTT	574
Qy	1081	GGGCGCTTTGGCGCGTGCGCGCGCTTCGCACTTCGGGGGTCATTATCACTAGACGGGA	1140
Db	575	GGGTTTTTCGGGTCGTTGTCGTA GTTTTCGTA TTTCGGGGGTTTATTATTAGACGGG	634
Qy	1141	CGGCCCGGGTG	1150
Db	635	CGGTCGGGTG	644
RESULT 7			
ABQ52277/c			
ID. ABQ52277 standard; DNA; 723 BP.			
XX	XX	ABQ52277;	
XX	XX		
DT	12-JUL-2002	(first entry)	
XX	XX		
XX	Oligonucleotide for detecting cytosine methylation SEQ ID NO 38868.		
KW	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;		
KW	drug; side effect; cancer; central nervous system; cardiovascular;		
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;		
KW	SNP; cell differentiation; ds.		
XX	Homo sapiens.		
OS	WO200218632-A2.		
PX	07-MAR-2002.		
PD	01-SEP-2001; 2001WO-BP10074.		
XX	01-SEP-2000; 2000DE-1043826.		
PR	05-SEP-2000; 2000DE-1044543.		
FR	{EPIG-} EPIGENOMICS AG.		
PA	Olek A, Piepenbrock C, Berlin K, Guetig D;		
XX	WPI; 2002-371829/40.		
DR	Determining the degree of cytosine methylation in genomic DNA, useful		
PT	for diagnosis and prognosis, comprises selective hybridization of		
PT	amplicons from chemically treated DNA		

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX Sequence 723 BP; 306 A; 229 C; 89 G; 99 T; 0 other;

Query Match 15.0%; Score 262.8; DB 24; Length 723;
 Best Local Similarity 65.8%; Pred. No. 1.1e-42;
 Matches 400; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 544 AAGGCGGCGGCGAGATGTCGAGTTGTGGCAGTTTCTACTGAGCTGTGGCAGACCGCGG 603
 DB 689 ATAGGTAGCGGATAGATTAGTTGTGTTGTTTGTGGAGTTGTGTTGATCGCGG 630
 QY 604 AACGCCGTGTCATCCGCTGGGAGCGCGCCAGCGAGTTTCAAGCTCACCGACCCGAC 663
 DB 629 AACGTCGGTTGTATCGCTGGGAGCGGTTACGCGAGTTTAAAGTTTACGGATTCCGAC 570
 QY 664 GAGTGGCGCGAGCTGGGCGGAGCGCAAGAGCAAGCCCAATATGAACTACGACAGCTA 723
 DB 569 GAGTGGCGCGGCGGTGGGCGGAGCGTAAGAGTAAGTTTAAATATGAATACGATAAGTTG 510
 QY 724 AGTCGAGCACTGCGCTACTACTACGACAAACATCATGAGCAAGGTGCGACGCGAGCGC 783
 DB 509 AGTCGCGTTTGGCTTATTATACGATAGATATATATGATAGGTGTATGTAAGCGT 450
 QY 784 TAGCGCTACCGTTTCACTTCCAGGCGCTGGACAGCTTGCAGCCACACCGCGCAC 843
 DB 449 TAGCTTTATCGTTTCGATTTTGGGTTTGGCGTAGGTTTGTAGTCGTTCGTCGTAC 390
 QY 844 GCCCAGCGCGCTCCGCGCGCGGAGCGGAGCGCGCCAGGATCGCGACTTTAC 903
 DB 389 GTTATGTCGTCGTAGTTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 330
 QY 904 AAGCTCCGCGCTGCTGGCTCCACTGCGCTTCCCGCGCTCTCCAAACTCAACCTTATG 963
 DB 329 AAGTTGTCGTCGTTTGGTTGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 270
 QY 964 GCAGCCTCGCGCGCTGGCGCGCTGGCTTCTTATCTGGCTGGTCC---CAACGCC 1020
 DB 269 GTCGTTTCGCTGGGCTGCGTTTCGTCGCTTTTATTTGTCGCGGTTTCGCTTCGTC 210
 QY 1021 ACCGCGCTGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 DB 209 GTTATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 150
 QY 1081 GGGCCCTTTGGCGCGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 DB 149 GGGTTTTTCGGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 90
 QY 1141 CGGCGGGGTG 1150
 DB 89 CGGTCGGGTG 80

RESULT 8

AAAX29137/c
 ID AAX29137 standard; cDNA; 1754 BP.
 AC AAX29137;
 XX
 DT 04-JUN-1999 (first entry)
 XX
 DE Hypoxia-regulated gene sequence RTP801.
 XX
 KW Hypoxia-regulated gene; therapeutic; diagnostic; hypoxia; ischemia;
 KW apoptosis; angiogenesis; tumorigenic cell; trauma; limb reattachment;
 KW revascularisation; ss.
 XX
 OS Rattus sp.
 XX
 PN WO9909049-A1.
 XX
 PD 25-FEB-1999.
 XX
 PF 21-AUG-1998; 98WO-US17296.
 XX
 PR 21-AUG-1997; 97US-0056453.
 XX
 PA (KOHN/) KOHN K I.
 PA (QUAR-) QUARK BIOTECH INC.
 XX
 PI Binat P, Skaliter R;
 XX
 DR WPI; 1999-180965/15.
 DR P-PSDB; AAY03634.
 XX
 PT New isolated hypoxia-related genes - used to develop products for
 PT use in therapy and diagnosis in e.g. hypoxia, ischemia, apoptosis
 PT and angiogenesis
 XX
 PS Claim 1; Page 69-70; 92pp; English.
 XX
 CC Sequences AAX29137 to AAX29142 represent isolated hypoxia-regulated
 CC genes. The genes and their products can be therapeutically and
 CC diagnostically used in hypoxia, ischemia, apoptosis and angiogenesis. The
 CC products and methods can be used for e.g. inducing apoptosis in
 CC tumorigenic cells or angiogenesis in trauma situations where e.g. a limb
 CC must be reattached or in a transplant where revascularisation is needed.
 XX
 SQ Sequence 1754 BP; 356 A; 459 C; 546 G; 393 T; 0 other;

Query Match 13.3%; Score 233.4; DB 20; Length 1754;
 Best Local Similarity 99.6%; Pred. No. 8.6e-37;
 Matches 234; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTTGTTTAAACAAACATGTTTATTAGAAAAGTAAATAATTCATAGGCTTAGTACTTG 60
 DB 1733 TTTGTTTAAACAAACATGTTTATTAGAAAAGTAAATAATTCATAGGCTTAGTACTTG 1674
 QY 61 AACATCAAGTGTTTATCATGAACCGTGAGTATCTTCATGTAAACAGTCTTAGATGGAAGAC 120
 DB 1673 AACATCAAGTGTTTATCATGAACCGTGAGTATCTTCATGTAAACAGTCTTAGATGGAAGAC 1614
 QY 121 CCAGTGGCGCTCTCTGGGGAGAGGGTTCCAGCCCCCCCCACCCCTCAGCCCATCCC 180
 DB 1613 CCAGTGGCGCTCTCTGGGGAGAGGGTTCCAGCCCCCCCCACCCCTCAGCCCATCCC 1554
 QY 181 CTCACAGTCTACTCTCTCCAGTACACCGCACCGGATGGCTGGGATGCGAGTCC 235
 DB 1553 CTCACAGTCTACTCTCTCCAGTACACCGCACCGGATGGCTGGGATGCGAGTCC 1499

RESULT 9

ABQ25880/c
 ID ABQ25880 standard; DNA; 719 BP.

Claim 12: 56pp + Sequence Listing: 56pp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 719 BP; 255 A; 268 C; 83 G; 113 T; 0 other;

Query Match	12.7%	Score 223;	DB 24;	Length 719;
Best Local Similarity	61.2%	Prod. No. 8e-35;		
Matches 413;	Conservative 0;	Mismatches 255;	Indels 7;	Gaps 3
Qy	609	CGGCTGCATCGCTGGGAGGCGGCACGGCGAGTTCAAGCTCACGACCCCGACGAGGT	668	
Db	9	CGACTACATCGCTAAAAAACGATCAGACGAATTCAACTCAGAACCCGACGAAT	68	
Qy	669	GGCGCGAGCTGGGGGAGCGCAGAGCAGGCCAATATGAATACGACAAAGTAAAGTCG	728	
Db	69	AACGCGACGATAAAACGACGCAAAAAAACAACCCCAACATAAATACTACGACAACTAAACCG	128	
Qy	729	AGCACTCGCTACTACTACGACAAAAACATCATGAGCAAGGTGACGCGCAAGCGTACGC	788	
Db	129	CGCCTACTAT-CTACTACGACAAAAACATCATAAACAAATACATAATAACGCTACGC	187	
Qy	789	CTACCGTTTGACTTCCAGGSCCTGGCAGAGCTTGCAGGCCACACCGCGCAGCCCA	848	
Db	188	CTACCGTTTGACTTCCAAAACCTAAACGCAAAATCTACCAACCGCGCGCGGACGCTCA	247	
Qy	849	CGCGCGCTGCGCGCGCGCAGCGGACGCGCCGCCAGATGCGGCACITTTACAAGT	908	
Db	248	TACCGCGCGCGCACTACTACGCGCGCG-CTGCGCCAAACGACGCGCTCTCAAACT	306	
Qy	909	CCGGCTGGTCTGGCTCCACTGCCCTTCCCGGGCTCTCCAAACTCAACCTTATGGCAGC	968	
Db	307	ACCGCGCAGCTCGACCGCTACGCTTCCCGCAGCCTCTCCAAACTCAACCTCATAACCGC	366	
Qy	969	CTGCGCGCGTGGCGCGCTGGCTCTCTTACTTGGCTGGTCCCAAGGCACCGCGCG	1026	
Db	367	CTCGCTAAAAATCGCGCGGACGACTTCTCTACTAAACCGAACCGATACCGACGCCACCGC	426	
Qy	1029	TGCGCGCGCACCGCTGCGCTCTACCCAAACCCGGGCTTGAGCGCCCTCCCGGGCCCTT	1088	
Db	427	TAAAAACGCCAAACCGCGCTCTAACCCCAATCGCAACTTACAAACCCCAACCAACGACAT	486	
Qy	1089	TGGCGCGGTGGCGCGCTTGGCACTTGGGGGGTCAATTATCACTAGACGGACGCGCGGG	1148	
Db	487	CGACCGTAAACGTCAACCTCGCACTTAAAAAACCATTTACCACTAAACAAAACGATCGAA	546	
Qy	1149	TGC-----AGTGGGGCTCTCCCAACAGCCAGTGACCAATCCCATCTCTCTCTGGGAG	1203	
Db	547	TACCTACGACTCGACCGAAGCTCTTAAATCTCGCGTATCCCATCTACATCTCGAAA	606	
Qy	1204	GAGCCCGAAGATTTCCCGACGTTTCTTTACACAGATTTGTTGACGAGCGCTCCC	1263	
Db	607	AATTCGGAAACCTTCGTCGACCGCTCTTAATCCAAATTTACTTCACTACCTACCGCACTT	666	

Query Match	12.78;	Score 223;	DB 22;	Length 2532;
Best Local Similarity	72.23;	Pred. No. 1.1e-34;		
Matches 366;	Conservative 0;	Mismatches 70;	Indels 71;	Gaps 3;
Qy	546	AGGCAGCGGGCAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGCAGACCGCGCGAA	605	
Db	2096	ATGCAGCGGACAGATCCAGCTGTGGCAGTTTCTGCTGGAGCTGCTGCCTGACCGCGCGAA	2155	
Qy	606	CGCGCGGTGCATCGCTGTGGAGGCGGCCACGGCGAGTTCAAGCTCACCGACCCCGGACGA	665	
Db	2156	CG-CGGCTGCATCGCTGTGGAGGCGGTCACGGCGAGTTCAAGCTCACCGACCCCGGACGA	2214	
Qy	666	GGTGGCGGACGCTGGGGCGAGCGCAGAGCAAGCCCAATATGAACCTACGACAGCTTAAG	725	
Db	2215	GGTGGCGGCGGTGGGGCGAGCGCAAGCAAGCCCAATGAATACGACAGCTGAG	2274	
Qy	726	TCGAGCACTGGCGTACTACTACGACAAAAACATCATGAGCAAGGTGCACGGCAAGCGCTA	785	
Db	2275	CCGGCGCCTGCTAT-CTACTACGACAGAACATCATGACAGGTGCATGGTAGCGCTA	2333	
Qy	786	CGCTACCGCTTTGACTTCCAGGCGCTTGGCAGAGGCTTGGCAGCACCCCGCGCAGCG	845	
Db	2334	CGCCTACCGCTTCGACTTCCAGGCGCTG-----	2361	
Qy	846	CCACCGCGCGCTCCGCGCGCGAGCGGCAGCGGCGCGCTTCCAGCAAGTGGCGCATTTACAA	905	
Db	2362	-----GCGCAGGACGGCGGCTCTACAA	2384	
Qy	906	GCTCCCGCTGGTGTGCTCCACTGCCCTTCCCGGCGCTTCCAAACTCAACCTTATGGC	965	
Db	2385	GCTGCCCGCGGCTTCGACCGCTGCGCTTCCCGCGGCTTCCAAACTCAACCTCATGGC	2444	
Qy	966	AGCTCGCGCGCGTGGCGCGCGCTTGCGTTCTCTTACTGGCGCTGGTCCCAACGCCACCGC	1025	
Db	2445	CGCTCGCTGGAGTTCGGCGCGACGGCTTCTCTACTTGGCGGCGCGGTGCCGACGCCAC	2504	
Qy	1026	CGCTGCCCGCGCCACCGCTGGCGTCTTA	1052	
Db	2505	CGCTGAAGCGCGCAGCGCGCGCTCTA	2531	

RESULT 12	
ABZ68768	
ID	ABZ68768 standard; DNA; 1219 BP.
XX	
XX	
AC	ABZ68768;
XX	
DT	16-MAY-2003 (first entry)
XX	
DE	Nucleotide sequence of human ERG splice variant C-1-3.
XX	
KW	Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
KW	C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
KW	cartilage disease; tumour; gene; ss.
XX	
XX	
OS	Homo sapiens.
XX	
XX	
Key	Location/Qualifiers
FT	146..1156
CDS	
FT	/*tag= a
FT	/product= "ERG splice variant C-1-3"
XX	
XX	
PN	WO2003000724-A2.
XX	
PD	03-JAN-2003.
XX	
XX	08-MAR-2002; 2002WO-EP02605.
PF	
XX	
PR	25-JUN-2001; 2001US-300756P.
XX	
XX	
PA	(TIGE-) TIGENIX NV.
XX	
XX	
PI	Luyten F, De Bari C, Dell'Accio F;

WPI; 2003-221398/21.
P-PSDB; ABP97697 .

Novel splice variants of the Ets-related gene (ERG) transcription factor , useful for preventing or treating bone-related or cartilage-related diseases or Ets-related gene associated tumour -

Claim 11; Fig 6; 60pp; English.

The present sequence encodes splice variant C-1-3 of human ERG. ERG belongs to the Ets family of transcription factors. The specification describes C-1-1 (characterised by the absence of the 81 bp fragment, and the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by the absence of the 218 bp fragment, and the presence of the 72 bp and 81 bp fragments of the ERG-3 cDNA sequence). The splice variant polynucleotides are useful for stabilising the phenotype of chondrocytes or cells with a chondrocytes-like phenotype, for preventing de-differentiation of articular chondrocytes during in vitro culture expansion, for preventing or treating bone-related or cartilage-related diseases or Ets-related gene associated tumour, or as marker of progenitor cells of stable chondrocytes. Cells expressing the polynucleotide are useful for producing or repairing cartilage tissue in a mammal.

SQ	Sequence	1219 BP; 312 A; 393 C; 288 G; 226 T; 0 other;
	Query Match	11.4%; Score 200.4; DB 25; Length 1219;
	Best Local Similarity	77.4%; Pred. No. 2.7e-30;
	Matches 24;	Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy	529	AGCCCTCGGTTACAGAAAGGCAGGGCGCAGATCCAGTTGTGGCAGTTCCTACTGGAGCTG 588
Db	617	AGCGCCCTTGCNAATCCAGGCACTGGCGCATCCAGCTTTGGCAGTTCTCTCTGGAGTC 676
Qy	589	CTGSCAGACC CGCGGAACGCCGGTGTCATCGCTGGGAGGCGCCACGGCGAGTTCCAAG 648
Db	677	CTGTGGGACAGCTCCAACTCCAGCTGCATCACCTGGGAGGCCAACCGGGGAGTTCCAAG 736
Qy	649	CTCACCAGCCCCCGACGAGGTGGCGCAGCTGGGGCGAGCGCAAGACAAGCCCCAATATG 708
Db	737	ATGACGGATCCCAGCAGAGTGGCCCGCGCTGGGAGAGCGGAGAGCAACCCCAACATG 796
Qy	709	AATACGACAAAGCTTAAGTTCGAGCACTCGGCTACTACTACGACAAAAACATCATGAGCAAG 768
Db	797	AACTACGATAAGCTTCAGCCGCGCCCTCCGTTACTACTATGACAAGAACATCATGACCAAG 856
Qy	769	GTGCACGGCAAGCGCTACGCCTTACCGCTTTGACCTTCAGGGCGCTGGGCACAGGCTTGCAG 828
Db	857	GTCATGGGAAGCGCTACGCCTCAAAAGTTCGACTTCCACGGGATCGCCCGCCCTCCAG 916
Qy	829	CCACCACCCGGCA 842
Db	917	CCCCACCCCGGGA 930

RESULT 13	
ABZ68767	
ID	ABZ68767 standard; DNA; 1291 BP.
XX	
XX	
AC	ABZ68767;
XX	
DT	16-MAY-2003 (first entry)
XX	
DE	Nucleotide sequence of human ERG splice variant C-1-2.
XX	
XX	Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
KW	C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
KW	cartilage disease; tumour; gene; ss.

```
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 146..1228
XX FT /*tag= a
XX FT /*product= "ERG splice variant C-1-2"
XX
XX WO2003000724-A2.
XX
XX PD 03-JAN-2003.
XX
XX PF 08-MAR-2002; 2002WO-EP02605.
XX
XX PR 25-JUN-2001; 2001US-300756P.
XX
XX PA (TIGEX-) TIGENIX NV.
XX
XX PI Luyten F, De Bari C, Dell'Accio F;
XX WPI: 2003-221398/21.
XX P-PSDB; ABP97696.
XX
XX Novel splice variants of the Ets-related gene (ERG) transcription
XX factor , useful for preventing or treating bone-related or
XX cartilage-related diseases or Ets-related gene associated tumour
XX
XX Claim 10; Fig 5; 60pp; English.
XX
XX The present sequence encodes splice variant C-1-2 of human ERG. ERG
XX belongs to the Ets family of transcription factors. The specification
XX describes C-1-1 (characterised by the absence of the 81 bp fragment, and
XX the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA
XX sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp
XX fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA
XX sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and
XX 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by
XX the absence of the 218 bp fragment, and the presence of the 72 bp and
XX 81 bp fragments of the ERG-3 cDNA sequence). The splice variant
XX polynucleotides are useful for stabilizing the phenotype of chondrocytes
XX or cells with a chondrocytes-like phenotype, for preventing
XX de-differentiation of articular chondrocytes during in vitro culture
XX expansion, for preventing or treating bone-related or cartilage-related
XX diseases or Ets-related gene associated tumour, or as marker of
XX progenitor cells of stable chondrocytes. Cells expressing the
XX polynucleotide are useful for producing or repairing cartilage tissue
XX in a mammal.
XX
XX Sequence 1291 BP; 334 A; 409 C; 302 G; 246 T; 0 other;
XX
XX Query Match 11.4%; Score 200.4; DB 25; Length 1291;
XX Best Local Similarity 77.4%; Pred. No. 2.7e-30;
XX Matches 243; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
XX
XX QY 529 AGCCCTCGGTTACAGAAAGGCGGCGGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTG 588
XX Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 689 AGCCGCTTGCAATCCAGGAGTGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGCTC 748
XX
XX QY 589 CTGGCAGACCGCGCAACGCGGCTGTCATCGGTGGAGGGGGCGGCGCAGCGGAGTTCAAG 648
XX Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 749 CTGTGGGACAGCTCCAACTCCAGCTGCATCACCTGGGAAGGACCACCGGGAGTTCAAG 808
XX
XX QY 649 CTCACGACCCCGAGAGTGCGCGGACGCTGGGCGGAGCGCAAGCAAGCCCAATATG 708
XX Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 809 ATGACGGATCCGACGAGGTGCCCGCGCTGGGAGAGCGGAGAGCAACCCCAACATG 868
XX
XX QY 709 AACTACGACAGCTAAGTTCGACACTGGCGTACTACTACGACAAACATCATGACCAAG 768
XX Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 869 AACTACGATAAGCTCAGCGCGCCCTCCGTTACTACTATGACAAACATCATGACCAAG 928
XX
XX QY 769 GTGCAGCGCAGCGCTAGCGCTTACGCTTTGACTTCCAGGGCCTGGGCAAGCTTGCCAG 828
XX Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 929 GTCCATGGGAAGCGGTACGCTTACAAAGTTCGACTTCCACGGGATCGCCCGAGCCCTCCAG 988
```

```
QY 829 CCACCACCGCGGCA 842
Db ||||| ||||| |||||
XX 989 CCCACCCCCCGGA 1002
XX
XX RESULT 14
XX ABZ68769
XX ID ABZ68769 standard; DNA; 1372 BP.
XX AC ABZ68769;
XX
XX DT 16-MAY-2003 (first entry)
XX
XX DE Nucleotide sequence of human ERG splice variant C-1-4.
XX
XX KW Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
XX C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
XX cartilage disease; tumour; gene; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX CDS 146..1309
XX FT /*tag= a
XX FT /*product= "ERG splice variant C-1-4"
XX
XX PN WO2003000724-A2.
XX
XX PD 03-JAN-2003.
XX
XX PF 08-MAR-2002; 2002WO-EP02605.
XX
XX PR 25-JUN-2001; 2001US-300756P.
XX
XX PA (TIGEX-) TIGENIX NV.
XX
XX PI Luyten F, De Bari C, Dell'Accio F;
XX WPI: 2003-221398/21.
XX P-PSDB; ABP97698.
XX
XX Novel splice variants of the Ets-related gene (ERG) transcription
XX factor , useful for preventing or treating bone-related or
XX cartilage-related diseases or Ets-related gene associated tumour
XX
XX Claim 112; Fig 7; 60pp; English.
```

```
XX The present sequence encodes splice variant C-1-4 of human ERG. ERG
XX belongs to the Ets family of transcription factors. The specification
XX describes C-1-1 (characterised by the absence of the 81 bp fragment, and
XX the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA
XX sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp
XX fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA
XX sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and
XX 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by
XX the absence of the 218 bp fragment, and the presence of the 72 bp and
XX 81 bp fragments of the ERG-3 cDNA sequence). The splice variant
XX polynucleotides are useful for stabilizing the phenotype of chondrocytes
XX or cells with a chondrocytes-like phenotype, for preventing
XX de-differentiation of articular chondrocytes during in vitro culture
XX expansion, for preventing or treating bone-related or cartilage-related
XX diseases or Ets-related gene associated tumour, or as marker of
XX progenitor cells of stable chondrocytes. Cells expressing the
XX polynucleotide are useful for producing or repairing cartilage tissue
XX in a mammal.
```

Sequence 1372 BP; 359 A; 430 C; 314 G; 269 T; 0 other;

Query Match 11.4%; Score 200.4; DB 25; Length 1372;
Best Local Similarity 77.4%; Pred. No. 2.7e-30;
Matches 243; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
QY 529 AGCCCTCGGTACAGAAAGGCGGCGGAGATCCAGTTGTGGCAGTTTCTACTGAGCTG 588
    |||||
Db 770 AGCCGCTTGCATAATCCAGGAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCGAGCTC 829
    |||||
QY 589 CTGGCAGACCGCGAAGCGCGGTGTCATCGGTGGGAGGGCGGCCACGGCGAGTTCAAG 648
    |||||
Db 830 CTGTGGACAGCTCCACTCCAGCTGTCATCAGCTGGGAGGACCACCAACGGGAGTTCAAG 889
    |||||
QY 649 CTCACCGACCGGACGAGGTGGCGGACGCTGGGCGGAGCGCAAGCAAGCCCAATATG 708
    |||||
Db 890 ATGACGGATCCGACGAGGTGGCCGCGCTGGGCGAGAGCGCAAGCAACCCCAACATG 949
    |||||
QY 709 AACTACGACAGCTTAAGTCGAGCATCGCTACTACTAGCAGCAAAAACATCATGAGCAAG 768
    |||||
Db 950 AACTACGATAAGCTCAGCGCGCCCTCCGTTACTACTATGACAAAGAACATCATGACCAAG 1009
    |||||
QY 769 GTGACCGCAAGCGCTACGCTTACCGCTTTCAGTTTCCAGGGCTTGGCAGAGCTTGGCCAG 828
    |||||
Db 1010 GTCCATGGGAGGCTAGCGCTACAGTTTCAGTTTCAGGGATCGCCAGGCCCTCCAG 1069
    |||||
QY 829 CCACACCGCGCA 842
    |||||
Db 1070 CCCCACCCCCCGGA 1083
    |||||
```

RESULT 15

ID ABX08759 standard; cDNA; 1389 BP.

AC ABX08759;

XX 21-JAN-2003 (first entry)

DE Angiogenesis-associated human polynucleotide sequence #21.

XX Human; angiogenesis-associated transcript; angiogenesis;

KW angiogenesis-associated disease; cancer; cytostatic; gene therapy;

KW gene; ss.

XX Homo sapiens.

OS WO200279492-A2.

XX 10-OCT-2002.

XX 14-FEB-2002; 2002WO-US04915.

XX 14-FEB-2001; 2001US-0784356.

PR 22-FEB-2001; 2001US-0791390.

PR 19-APR-2001; 2001US-285475P.

PR 03-AUG-2001; 2001US-310025P.

PR 13-NOV-2001; 2001US-350666P.

PR 29-NOV-2001; 2001US-334244P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Murray R, Glynn R, Watson SR, Aziz N;

XX WPI; 2003-040681/03.

XX P-PSDB; ABU03476.

XX Detecting angiogenesis-associated transcript in a cell for diagnosing
PT and treating cancer by contacting a sample with a polynucleotide that
PT exhibits changes in expression level as a function of time in tissue
PT undergoing angiogenesis -

PS Example 2; Page 200; 291pp; English.

XX The present invention relates to methods and compositions for

CC detecting an angiogenesis-associated transcript in a cell in

CC a patient. The method involves contacting a biological sample from

CC the patient with a polynucleotide that selectively hybridizes to a

CC sequence at least 80% identical to any of the angiogenesis-associated

CC human polynucleotide sequences given in the specification. These
CC angiogenesis-associated polynucleotide sequences comprise genes that
CC exhibit changes in expression levels as a function of time in tissue
CC undergoing angiogenesis. The method and the polynucleotide sequences
CC of the invention are useful for diagnosing and treating angiogenesis
CC and angiogenesis-associated diseases e.g. cancer. The polynucleotide
CC sequences are also useful in the gene therapy of such disorders. The
CC angiogenesis-associated proteins encoded by the polynucleotide
CC sequences are useful as a vaccine for therapeutic and prophylactic
CC immunisation. ABX08739-ABX08853 represent angiogenesis-associated
CC polynucleotide sequences.

XX Sequence 1389 BP; 359 A; 440 C; 323 G; 267 T; 0 other;

Query Match 11.4%; Score 200.4; DB 25; Length 1389;

Best Local Similarity 77.4%; Pred. No. 2.7e-30;

Matches 243; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 529 AGCCCTCGGTACAGAAAGGCGGCGGAGATCCAGTTGTGGCAGTTTCTACTGAGCTG 588

Db 850 AGCCGCTTGCATAATCCAGGAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCGAGCTC 909

QY 589 CTGGCAGACCGCGAAGCGCGGTGTCATCGGTGGGAGGGCGGCCACGGCGAGTTCAAG 648

Db 910 CTGTGGACAGCTCCAACTCCAGCTGCATCACCCTGGGAAAGGCACCAACGGGAGTTCAAG 969

QY 649 CTCACCGACCGGACGAGGTGGCGGACGCTGGGCGGAGCGCAAGCAAGCCCAATATG 708

Db 970 ATGACGGATCCCGACGAGGTGGCCGCGCTGGGAGAGCGGAGCAACCCCAACATG 1029

QY 709 AACTACGACAGCTTAAGTCGAGCACTGCGCTACTACTACGACAAAACATCATGAGCAAG 768

Db 1030 AACTACGATAAGCTCAGCGCGCTCCGTTACTACTATGACAAAGAACATCATGACCAAG 1089

QY 769 GTGACCGCAAGCGCTACGCTTTCAGTTTCCAGGGCTTGGCAGAGCTTGGCCAG 828

Db 1090 GTCCATGGGAGGCTAGCGCTACAGTTTCAGTTTCAGGGATCGCCAGGCCCTCCAG 1149

QY 829 CCACACCGCGCA 842

Db 1150 CCCCACCCCCCGGA 1163

Search completed: November 26, 2003, 13:33:03

Job time : 495 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 13:24:47 ; Search time 571 Seconds
(without alignments)
10109.102 Million cell updates/sec

Title: US-10-027-859-1

Perfecc score: 1752
Sequence: 1 ttgttttaacaaacattgtt.....aaaaaaaaaaaaaaaaaaaaa 1752

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PTCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1752	100.0	1752	US-09-850-799-1	Sequence 1, Appli
2	1752	100.0	1752	US-10-027-859-1	Sequence 1, Appli
C 3	233.4	13.3	1754	US-10-091-333-1	Sequence 1, Appli
C 4	233.4	13.3	1754	US-10-325-878-1	Sequence 1, Appli
5	223	12.7	2532	US-10-037-270-869	Sequence 869, App
6	200.4	11.4	3166	US-10-021-660-261	Sequence 26, Appl
7	200.4	11.4	3166	US-10-205-823-98	Sequence 98, Appl
C 8	199.4	11.4	567	US-09-864-761-18410	Sequence 18410, A
C 9	198.4	11.3	473	US-09-864-761-20472	Sequence 20472, A
C 10	197.4	11.3	472	US-10-029-386-25224	Sequence 25224, A
11	194	11.1	2957	US-10-007-926A-295	Sequence 295, App
12	180	10.3	1447	US-09-902-772-1	Sequence 1, Appli
13	180	10.3	1528	US-09-902-772-3	Sequence 3, Appli
14	138.8	7.9	441	US-10-210-120-41	Sequence 41, Appl
15	138.8	7.9	473	US-09-918-995-27904	Sequence 27904, A
16	138.8	7.9	1884	US-09-925-300-420	Sequence 420, App

17	138.8	7.9	2188	10	US-09-920-300A-1716	Sequence 1716, Ap
18	138.8	7.9	2188	12	US-10-099-926-1716	Sequence 1716, Ap
19	138.8	7.9	2188	13	US-10-033-528-1716	Sequence 1716, Ap
20	138.8	7.9	2268	10	US-09-920-300A-1693	Sequence 1693, Ap
21	138.8	7.9	2268	12	US-10-099-926-1693	Sequence 1693, Ap
22	138.8	7.9	2268	13	US-10-033-528-1693	Sequence 1693, Ap
23	138.8	7.9	2269	10	US-09-954-531-955	Sequence 955, App
24	138.8	7.9	3692	14	US-10-106-698-824	Sequence 824, App
C 25	136.2	7.8	530	12	US-10-029-386-11517	Sequence 11517, A
26	132.2	7.5	3178	13	US-10-108-605-124	Sequence 124, App
27	132.2	7.5	3178	13	US-10-108-605-128	Sequence 128, App
28	125.4	7.2	549	9	US-09-923-779-54	Sequence 54, Appl
C 29	121	6.9	703	14	US-10-050-704-18	Sequence 18, Appl
C 30	121	6.9	1734	9	US-09-739-262-5	Sequence 5, Appli
C 31	121	6.9	1744	10	US-09-919-172-45	Sequence 45, Appl
C 32	121	6.9	1744	11	US-09-919-039-330	Sequence 330, App
C 33	121	6.9	1744	12	US-10-051-835-23	Sequence 23, Appl
C 34	121	6.9	1760	12	US-10-301-822-60	Sequence 60, Appl
C 35	121	6.9	1760	12	US-10-465-572-5	Sequence 5, Appli
C 36	121	6.9	1760	14	US-10-171-311-57	Sequence 57, Appl
C 37	121	6.9	1760	14	US-10-175-533-173	Sequence 173, App
C 38	121	6.9	1782	14	US-10-091-333-2	Sequence 2, Appli
C 39	121	6.9	1782	14	US-10-325-878-2	Sequence 2, Appli
C 40	119.4	6.8	2142	12	US-09-814-353-20061	Sequence 20061, A
C 41	119.4	6.8	1023	13	US-10-002-600-2	Sequence 2, Appli
C 42	118	6.7	366	10	US-09-880-107-999	Sequence 999, App
C 43	118	6.7	366	12	US-09-873-319-188	Sequence 188, App
C 44	118	6.7	366	12	US-09-960-706-318	Sequence 318, App
C 45	118	6.7	366	12	US-09-873-367C-386	Sequence 386, App

ALIGNMENTS

RESULT 1

US-09-850-799-1
; Sequence 1, Application US/09850799
; Patent No. US20020030647A1
; GENERAL INFORMATION:
; APPLICANT: Deneis, Evan S.
; APPLICANT: Fyodor, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
; FILE REFERENCE: CASE-03828
; CURRENT APPLICATION NUMBER: US/09/850,799
; CURRENT FILING DATE: 2001-05-08 -
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/360,779
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-07-26
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(1131)
US-09-850-799-1

Query Match	100.0%	Score 1752;	DB 9;	Length 1752;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1752;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	TTTGTTTTAAACAACATGTTTATTAGAAAAGTAAATAATTGCATAGGCTTTAGTACTTG	60	
Db	1	TTTGTTTTAAACAACATGTTTATTAGAAAAGTAAATAATTGCATAGGCTTTAGTACTTG	60	
QY	61	AACATCAAGTGTATTTCATGAACCGTGACTTCTTTCATCTTAACACGTTCTAGATGAAGAC	120	
Db	61	AACATCAAGTGTATTTCATGAACCGTGACTTCTTTCATCTTAACACGTTCTAGATGAAGAC	120	
QY	121	CCAGGTGGCGTCTCTCTGGGGGAGAGGGTTCACAGCCCCCCCCCTCAGCCCCATCCC	180	

Db 121 CCAGGTGGCGCTCTCTGGGGAGAGGTTCCAGGCCCCCAGCCCTCAGCCCCATCCC 180
Qy 181 CTCACAGCTCACTCTCCAGTACACCGGCACCGGATGGCTGGGATGAGCTCCAGGAC 240
Db 181 CTCACAGCTCACTCTCCAGTACACCGGCACCGGATGGCTGGGATGAGCTCCAGGAC 240
Qy 241 CGCTCCCTCTCTACACAGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 300
Db 241 CGCTCCCTCTCTACACAGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 300
Qy 301 GGGTCTACTCTTCTCTCCACACAGTCCACACAGTCCACACAGTCCACACAGTCCACACAGT 360
Db 301 GGGTCTACTCTTCTCTCCACACAGTCCACACAGTCCACACAGTCCACACAGTCCACACAGT 360
Qy 361 AAGCTGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 420
Db 361 AAGCTGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 420
Qy 421 ATGAGACAGAGGACCTCCAGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 480
Db 421 ATGAGACAGAGGACCTCCAGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 480
Qy 481 GGAGATGGTCTTTTAAAGAAAGGAGAGCCGAGCTGGGGGCGCTGAGCCCTCGGTA 540
Db 481 GGAGATGGTCTTTTAAAGAAAGGAGAGCCGAGCTGGGGGCGCTGAGCCCTCGGTA 540
Qy 541 CAGAAAGGAGGCGGAGATCCAGTTGTGGAGTTTCTACTGGAGCTGCTGGCAGACCGC 600
Db 541 CAGAAAGGAGGCGGAGATCCAGTTGTGGAGTTTCTACTGGAGCTGCTGGCAGACCGC 600
Qy 601 GCGAAGCGGCTGATCCGCTGGAGGCGGCGGAGCTGGAGCTGGAGCTGGAGCTGGAGCT 660
Db 601 GCGAAGCGGCTGATCCGCTGGAGGCGGCGGAGCTGGAGCTGGAGCTGGAGCTGGAGCT 660
Qy 661 GACGAGGTGGCGGAGCTGGGCGGAGCGCAAGCAAGCCCAATATGAATACGACAAG 720
Db 661 GACGAGGTGGCGGAGCTGGGCGGAGCGCAAGCAAGCCCAATATGAATACGACAAG 720
Qy 721 CTAAGTCGAGCACTGGCTTACTACTACGACAAACATCATGAGCAAGGTGACGGCAAG 780
Db 721 CTAAGTCGAGCACTGGCTTACTACTACGACAAACATCATGAGCAAGGTGACGGCAAG 780
Qy 781 CGCTAGCCTACCGCTTTGACTTCAGGGCTGGCAGAGCTTGGCAGGCTTGGCAGGCTTGG 840
Db 781 CGCTAGCCTACCGCTTTGACTTCAGGGCTGGCAGAGCTTGGCAGGCTTGGCAGGCTTGG 840
Qy 841 CACGCCACGCGCGCTGCTT 900
Db 841 CACGCCACGCGCGCTGCTT 900
Qy 901 TACAAGCTCCGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 960
Db 901 TACAAGCTCCGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 960
Qy 961 ATGGAGCCTCGGCGGCTGGCGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1020
Db 961 ATGGAGCCTCGGCGGCTGGCGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1020
Qy 1021 ACCGCGCTGGCGCGCACCGCTGGCTGTACCAACCGCGGCTTGCAGCCCTCCCTCC 1080
Db 1021 ACCGCGCTGGCGCGCACCGCTGGCTGTACCAACCGCGGCTTGCAGCCCTCCCTCC 1080
Qy 1081 GGGCCCTTTGGCGGCTGGCGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCT 1140
Db 1081 GGGCCCTTTGGCGGCTGGCGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCT 1140
Qy 1141 CGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 1200
Db 1141 CGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 1200
Qy 1201 GAGGAGCCCGAAGATTTCCCGAGCTTCTTTTACACAGATTTTCGTTGCGAGCGCGCT 1260

Db 1201 GAGGAGCCCGAAGATTTCCCGAGCTTCCCTTTTACACAGATTTTCGTTGCGAGCGCGCT 1260
Qy 1261 CCCAGCCAGGGAAGAAAGGATGGGAAGCTCTGAGGCTTCTCTTGAATACGAGGCTTCC 1320
Db 1261 CCCAGCCAGGGAAGAAAGGATGGGAAGCTCTGAGGCTTCTCTTGAATACGAGGCTTCC 1320
Qy 1321 AGGCTCCCATATCATCATCCCCAGGAAGGTCATGTGCTCCCACTTTAAATTTTCTCTT 1380
Db 1321 AGGCTCCCATATCATCATCCCCAGGAAGGTCATGTGCTCCCACTTTAAATTTTCTCTT 1380
Qy 1381 CCAAGTCTCCAGATTTCTGMACTCCCGTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440
Db 1381 CCAAGTCTCCAGATTTCTGMACTCCCGTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440
Qy 1441 CTTCCTCTTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTCTC 1500
Db 1441 CTTCCTCTTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTCTC 1500
Qy 1501 ATTTTCTCTCTCCAGACTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGGCT 1560
Db 1501 ATTTTCTCTCTCCAGACTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGGCT 1560
Qy 1561 TCTCACACTCCCTTTTCCGGATATGAGAAGCATCAAAAACATCTCTGCTGTGTCATC 1620
Db 1561 TCTCACACTCCCTTTTCCGGATATGAGAAGCATCAAAAACATCTCTGCTGTGTCATC 1620
Qy 1621 CCTATCCCAACACTCTGCTTTCCTTCCATACCACTCTTGGCCCAAGGACCTCG 1680
Db 1621 CCTATCCCAACACTCTGCTTTCCTTCCATACCACTCTTGGCCCAAGGACCTCG 1680
Qy 1681 TCTGTATATATTCCTTTTCCAGCCCATTAAGATCCAAAGGATTTTCAAAAAA 1740
Db 1681 TCTGTATATATTCCTTTTCCAGCCCATTAAGATCCAAAGGATTTTCAAAAAA 1740
Qy 1741 AAAAAA 1752
Db 1741 AAAAAA 1752

RESULT 2
US-10-027-859-1
; Sequence 1, Application US/10027859
; Publication No. US20030175930A1
; GENERAL INFORMATION:
; APPLICANT: Deneris, Evan S.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
; FILE REFERENCE: CASE-04027
; CURRENT APPLICATION NUMBER: US/10/027,859
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US/09/435,335
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/360,779
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-07-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112) .. (1131)
US-10-027-859-1

Query Match 100.0%; Score 1752; DB 12; Length 1752;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTTGTTTAAACAAACATGTTTATTAGAAAAGTAAATAATTGATAGGCTTAGTACTTG 60

Db	1	TTTTGTTTAA	CAAA	CATGTTTATT	TAGAAAAGTAA	AAATAT	TGCA	TAGTCT	TAGTACT	TG	60
Qy	61	AACATCAAG	TGATAT	TATCA	TGAACCGT	GAGTATCTT	CATG	TAAACAGT	TCTTAGAT	GAAAGAC	120
Db	61	AACATCAAG	TGATAT	TATCA	TGAACCGT	GAGTATCTT	CATG	TAAACAGT	TCTTAGAT	GAAAGAC	120
Qy	121	CCAGTGGCG	TCTCTCT	GGGGAGAG	GGTTC	CACGCC	CCCA	CCCTC	TACG	CCCCCATCCC	180
Db	121	CCAGTGGCG	TCTCTCT	GGGGAGAG	GGTTC	CACGCC	CCCA	CCCTC	TACG	CCCCCATCCC	180
Qy	181	CTCACAGT	CTCACT	CTCT	CCAGTAC	ACCGGC	ACACGG	GATGGG	CTGGAT	GCAGCTCC	240
Db	181	CTCACAGT	CTCACT	CTCT	CCAGTAC	ACCGGC	ACACGG	GATGGG	CTGGAT	GCAGCTCC	240
Qy	241	CCCTCCCT	CTCTAC	CAACAC	CCCTGG	CTGCC	CTCC	CGC	CAGG	CTTGC	300
Db	241	CCCTCCCT	CTCTAC	CAACAC	CCCTGG	CTGCC	CTCC	CGC	CAGG	CTTGC	300
Qy	301	CGCTTACT	CTTCCCT	GTCA	CCCCAC	AGTCA	CAC	ACGG	GGCGGTAT	CGGCAC	360
Db	301	CGCTTACT	CTTCCCT	GTCA	CCCCAC	AGTCA	CAC	ACGG	GGCGGTAT	CGGCAC	360
Qy	361	AAGCTGAG	CTGCC	CCCCCGT	CGGT	TCCCC	ATCT	CCCC	ACCG	CCCCAGT	420
Db	361	AAGCTGAG	CTGCC	CCCCCGT	CGGT	TCCCC	ATCT	CCCC	ACCG	CCCCAGT	420
Qy	421	ATGAGACAG	GGCAG	CTCC	ACGCC	CTCT	GTGAT	CA	CACTG	TACCTAC	480
Db	421	ATGAGACAG	GGCAG	CTCC	ACGCC	CTCT	GTGAT	CA	CACTG	TACCTAC	480
Qy	481	GGAGATG	GTCTTTT	TAAG	GAAGGA	AGAC	CCCG	AGCT	GGGG	CGCGCT	540
Db	481	GGAGATG	GTCTTTT	TAAG	GAAGGA	AGAC	CCCG	AGCT	GGGG	CGCGCT	540
Qy	541	CAGAAAGC	AGCGG	CAGAT	CCAGT	TGTGG	CAGT	TTCT	ACTG	GAGCTG	600
Db	541	CAGAAAGC	AGCGG	CAGAT	CCAGT	TGTGG	CAGT	TTCT	ACTG	GAGCTG	600
Qy	601	CGCAAC	CGCGG	CTGC	TATCG	CTGG	AGGG	CGGCC	AC	CGCGAGT	660
Db	601	CGCAAC	CGCGG	CTGC	TATCG	CTGG	AGGG	CGGCC	AC	CGCGAGT	660
Qy	661	GACAGAGT	GGCG	CA	CGTGGG	CGAG	GGG	CAAG	CAAG	CCCAAT	720
Db	661	GACAGAGT	GGCG	CA	CGTGGG	CGAG	GGG	CAAG	CAAG	CCCAAT	720
Qy	721	CTAAGT	CGAG	CACT	CGCT	TACT	CTAC	GAC	AAAAA	CAAT	780
Db	721	CTAAGT	CGAG	CACT	CGCT	TACT	CTAC	GAC	AAAAA	CAAT	780
Qy	781	CGCTAC	CGCT	TAC	CGT	TGACT	TG	GCAC	AGG	CTTCC	840
Db	781	CGCTAC	CGCT	TAC	CGT	TGACT	TG	GCAC	AGG	CTTCC	840
Qy	841	CACGCC	CA	CGCG	CTG	CCG	CGC	AG	CGCG	CGCG	900
Db	841	CACGCC	CA	CGCG	CTG	CCG	CGC	AG	CGCG	CGCG	900
Qy	901	TACAAG	CTCC	CGG	CTG	GTG	CT	CG	CTT	CC	960
Db	901	TACAAG	CTCC	CGG	CTG	GTG	CT	CG	CTT	CC	960
Qy	961	ATGCGAG	CTCG	CGG	GTG	CG	CG	CTG	GTG	CTT	1020
Db	961	ATGCGAG	CTCG	CGG	GTG	CG	CG	CTG	GTG	CTT	1020
Qy	1021	ACCG	CGCT	TC	CG	CG	CA	CG	CT	T	1080
Db	1021	ACCG	CGCT	TC	CG	CG	CA	CG	CT	T	1080
Qy	1081	GGG	CCCTT	TG	CG	CGT	GG	CG	CT	TG	1140
Db	1081	GGG	CCCTT	TG	CG	CGT	GG	CG	CT	TG	1140

Qy	1141	CGGCCGGTGCAGTGGGGCCCTCTCCACACAGCCAGTGCACCAATCCCATCCTCATCTCTGG	1200
Db	1141	CGGCCGGTGCAGTGGGGCCCTCTCCACACAGCCAGTGCACCAATCCCATCCTCATCTCTGG	1200
Qy	1201	GAGGAGCCCGAAGATTTCCCGACGTTCTCTTTACACACAGATTTTCGTTGACAGCGCGCT	1260
Db	1201	GAGGAGCCCGAAGATTTCCCGACGTTCTCTTTACACACAGATTTTCGTTGACAGCGCGCT	1260
Qy	1261	CCACGCCACGGGAGAAAGGATGGGAAGCCTCTGAGGTCTTCTTGAATACGAGGTTCC	1320
Db	1261	CCACGCCACGGGAGAAAGGATGGGAAGCCTCTGAGGTCTTCTTGAATACGAGGTTCC	1320
Qy	1321	AGGCTCCCATATCATCACCCGAGGAAGGGTGATGTCTCCCACTTTAATTTTCTCTT	1380
Db	1321	AGGCTCCCATATCATCACCCGAGGAAGGGTGATGTCTCCCACTTTAATTTTCTCTT	1380
Qy	1381	CCAAGTCTCCAGATTTCTGGAACCTCCCGTCTTTTTTTTCTCTCACTCGAGGCCCTGC	1440
Db	1381	CCAAGTCTCCAGATTTCTGGAACCTCCCGTCTTTTTTCTCTCACTCGAGGCCCTGC	1440
Qy	1441	CTTCTCTTTATGACCCCTAGTTTTCTGTTTTGTTTTTTTTTTTTTCTCTCTCTCTCTC	1500
Db	1441	CTTCTCTTTATGACCCCTAGTTTTCTGTTTTGTTTTTTTTTTTTTCTCTCTCTCTCTC	1500
Qy	1501	ATTATTTTTCTCTCCACGACCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGGCT	1560
Db	1501	ATTATTTTTCTCTCCACGACCTACTCCAAACGGTAGTACCTCGGTAGTACCTCGAGGCT	1560
Qy	1561	TCTCACACTCCCTTTTTCGGGATATGAGAAGCATCAAAAACATCTCTGCTGTTGTCCATC	1620
Db	1561	TCTCACACTCCCTTTTTCGGGATATGAGAAGCATCAAAAACATCTCTGCTGTTGTCCATC	1620
Qy	1621	CCTATCCCAACACTCTGGCTTCGCTCCCTTCCATACACACTCTGGGCCCAAGGACCTCG	1680
Db	1621	CCTATCCCAACACTCTGGCTTCGCTCCCTTCCATACACACTCTGGGCCCAAGGACCTCG	1680
Qy	1681	TCTGTATATATCTCTTTTCAGCCCCATTTAAAGATCCCAAGCTTCAAAAAAATAAAAAA	1740
Db	1681	TCTGTATATATCTCTTTTCAGCCCCATTTAAAGATCCCAAGCTTCAAAAAAATAAAAAA	1740
Qy	1741	AAAAAAAAAAAAA 1752	
Db	1741	AAAAAAAAAAAAA 1752	

```

RESULT 3
US-10-091-333-1/c
; Sequence 1, Application US/10091333
; Publication No. US2003104973A1
; GENERAL INFORMATION:
; APPLICANT: EINAT, Paz
; APPLICANT: SKALITER, Rami
; TITLE OF INVENTION: HYPOXIA-REGULATED GENES
; FILE REFERENCE: EINAT=1.1D
; CURRENT APPLICATION NUMBER: US/10/091.333
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 09/604,978
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 09/138,112
; PRIOR FILING DATE: 1999-08-21
; PRIOR APPLICATION NUMBER: US 60/056,453
; PRIOR FILING DATE: 1997-08-21
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatenIn version 3.1
; SEQ ID NO 1
; LENGTH: 1754
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-091-333-1

```

Query Match 13.3%; Score 233.4; DB 14; Length 1754;
Best Local Similarity 99.6%; Pred. No. 2.8e-51;

Matches 234; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTTTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGTCATAGGCTTTAGTACTTG 60
Db 1733 TTGTTTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGTCATAGGCTTTAGTACTTG 1674

QY 61 AACATCAAGTGATTTCATGAACCGTGAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 120
Db 1673 AACATCAAGTGATTTCATGAACCGTGAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 1614

QY 121 CCAGGTGGCGCTCTCTGGGGAGAGGGTTCCAGCCCCCAGCCCCCTCAGCCCCATCCC 180
Db 1613 CCAGGTGGCGCTCTCTGGGGAGAGGGTTCCAGCCCCCAGCCCCCTCAGCCCCATCCC 1554

QY 181 CTCACAGCTCACTCTCCAGTACACCGGCACCGGATGGCTGGGATGCAGCTCC 235
Db 1553 CTCACAGTTCACTCTCCAGTACACCGGCACCGGATGGCTGGGATGCAGCTCC 1499

RESULT 4
US-10-325-878-1/c
; Sequence 1, Application US/10325878
; Publication No. US20030124116A1
; GENERAL INFORMATION:
; APPLICANT: Einat, Paz
; TITLE OF INVENTION: HYPOXIA-REGULATED GENES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 No. US20030124116A1althwestern Hwy., Suite 401
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/325,878
; FILING DATE: 23-Dec-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,112
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 0168,00034
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 5395055
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-325-878-1

Query Match 13.3%; Score 233.4; DB 14; Length 1754;
Best Local Similarity 99.6%; Pred. No. 2.8e-51;
Matches 234; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTTTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGTCATAGGCTTTAGTACTTG 60
Db 1733 TTGTTTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGTCATAGGCTTTAGTACTTG 1674

QY 61 AACATCAAGTGATTTCATGAACCGTGAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 120
Db 1673 AACATCAAGTGATTTCATGAACCGTGAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 1614

QY 121 CCAGGTGGCGCTCTCTGGGGAGAGGGTTCCAGCCCCCAGCCCCCTCAGCCCCATCCC 180
Db 1613 CCAGGTGGCGCTCTCTGGGGAGAGGGTTCCAGCCCCCAGCCCCCTCAGCCCCATCCC 1554

QY 181 CTCACAGCTCACTCTCCAGTACACCGGCACCGGATGGCTGGGATGCAGCTCC 235
Db 1553 CTCACAGTTCACTCTCCAGTACACCGGCACCGGATGGCTGGGATGCAGCTCC 1499

RESULT 5
US-10-037-270-869
; Sequence 869, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; PRIOR FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 869-
; LENGTH: 2532
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2532)
US-10-037-270-869

Query Match 12.7%; Score 223; DB 14; Length 2532;
Best Local Similarity 72.2%; Pred. No. 1.9e-48;
Matches 366; Conservative 0; Mismatches 70; Indels 71; Gaps 3;

QY 546 AGGCAGCGGCAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGCGCAA 605
Db 2096 ATGCAGCGGACAGATCCAGCTGTGCGTGTTCCTGTGGAGCTGCTGGCTGACCGCGCAA 2155

QY 605 CGCCGGCTGCATCGCTGGGAGCGGCACCGGAGTTCAAGCTCACCGACCCCGACCA 665
Db 2156 CG-CGGCTGCATCGCTGGGAGCGGTTCACCGCGAGTTCAAGCTCACCGACCGA 2214

QY 666 GTGCGGCGACCGCTGGGCGAGCGCAAGCAAGCCCAATATGAACCTACGAAGTAAG 725
Db 2215 GGTGCGGCGCGGTGGGCGAGCGCAAGCAAGCAAGCAAGCAAGCAAGCTGAG 2274

QY 726 TCGAGCACTGCGCTACTACTACGACAAAAACATCATGAGCAAGGTGCACGGCAGCGCTA 785

```
Db      2275  CGCGCCCTGCTAT-CTACTAGCAAGAACATCATGACGAAGGTGCATGTAAGCGCTA 2333
Qy      786  CGCCTACCGCTTTGACTTCCAGGGCTGGCACAGGTTGCCAGCCACACCGCGCACGC 845
Db      2334  CGCCTACCGCTTCGACTTCCAGGGCTG-----2361
Qy      846  CCACGCGCGCTGCGCGCGCGAGCGGACGCGCGCCAGGATGCGCATTTTACAA 905
Db      2362  -----GGCAGGACGGCGCTCTACAA 2384
Qy      906  GCTCCGGCTGCTGGCTCCACTGCTTCCCGGCTCTCCAACTCAACCTTATGCC 965
Db      2385  GCTGCCCGCGGCTCGACCGCTGCTTCCCGGCTCTCCAACTCAACCTATGCC 2444
Qy      966  AGCCTCGCGCGCTGGCGCGCTGCTTCTTACTGGCTGTGCTCCCAAGCCACCGC 1025
Db      2445  CGCTCGCTGGAGTGGCGCGCGGCTTCTCTACTGCGCGGCGGTGCGACGCCAC 2504
Qy      1026  CGCTCGCGCGCACCGCTGCGCTCTA 1052
Db      2505  CGCTGAAGCGCGCGCGCGCTCTA 2531
```

RESULT 6

```
US-10-021-660-26
; Sequence 26, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynné, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-0007100S
; CURRENT APPLICATION NUMBER: US/10/021,660
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 3166
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-021-660-26
```

```
Query Match      11.4%; Score 200.4; DB 12; Length 3166;
Best Local Similarity 77.4%; Pred. No. 2.1e-42;
Matches 243; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy      529  AGCCCTCGGTACAGAAAGGCGGCGAGATCCAGTTGTGGCACTTTCTACTGGAGCTG 588
Db      1106  AGCCGCTTGGAAATCCAGGAGTGGCCAGATCCAGCTTTGGCACTTCTCTGGAGCTC 1165
Qy      589  CTGGCAGACCGCGCAACCGCGGTGTCATCGCTGGGAGGCGGCCACCGGGAGTTCAAG 648
Db      1166  CTGTGGACAGCTCCAACTCCAGCTGTCATCACCTGGGAAGGCACCAACGGGAGTTCAAG 1225
Qy      649  CTCACCGACCCCGACGAGGTGGCGGAGCGCTGGGCGAGCGCAAGCAAGCCCAATATG 708
Db      1226  ATGACGGATCCGACAGGTGGCGCGCTGGGAGAGCGGAGAGCAACACCAACATG 1285
Qy      709  AACTACGACAGCTTAAGTCGAGCATCGGCTACTACTAGCAAAACATCATGAGCAAG 768
Db      1286  AACTACGATAAGCTCAGCGCGCTTACTACTATGACAAACATCATGACCAAG 1345
Qy      769  GTGCACGCGAAGCGCTACGCTTTCAGCTTTCAGGCGCTGGCACAGGCTTGCCAG 828
```

```
Db      1346  GTCCATGGGAAGCGCTACGCTTACAAAGTTCCAGCTTCCACGGGATCGCCACGGCCCTCCAG 1405
Qy      829  CCACCACCGCGCA 842
Db      1406  CCCACCCCCCGGA 1419

RESULT 7
US-10-205-823-98
; Sequence 98, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoerssch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsney, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 3166
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-98
```

```
Query Match      11.4%; Score 200.4; DB 14; Length 3166;
Best Local Similarity 77.4%; Pred. No. 2.1e-42;
Matches 243; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy      529  AGCCCTCGGTACAGAAAGGCGGCGAGATCCAGTTGTGGCACTTTCTACTGGAGCTG 588
Db      1106  AGCCGCTTGGAAATCCAGGAGTGGCCAGATCCAGCTTTGGCACTTCTCTGGAGCTC 1165
Qy      589  CTGGCAGACCGCGCAACCGCGGTGTCATCGCTGGGAGGCGGCCACCGCGAGTTCAAG 648
Db      1166  CTGTGGACAGCTCCAACTCCAGCTGTCATCACCTGGGAAGGCACCAACGGGAGTTCAAG 1225
Qy      649  CTCACCGACCCCGACGAGGTGGCGGAGCGCTGGGCGAGCGCAAGCAAGCCCAATATG 708
Db      1226  ATGACGGATCCGACAGGTGGCGCGCTGGGAGAGCGGAGAGCAACACCAACATG 1285
Qy      709  AACTACGACAGCTTAAGTCGAGCATCGGCTACTACTAGCAAAACATCATGAGCAAG 768
Db      1286  AACTACGATAAGCTCAGCGCGCTTCTACTATGACAAACATCATGACCAAG 1345
Qy      769  GTGCACGCGAAGCGCTACGCTTTCAGCTTTCAGGCGCTGGCACAGGCTTGCCAG 828
Db      1346  GTCCATGGGAAGCGCTACGCTTACAAAGTTCCAGCTTCCACGGGATCGCCACGGCCCTCCAG 1405
Qy      829  CCACCACCGCGCA 842
```

Db 1406 CCCACCCCCCGGA 1419

RESULT 8

US-09-864-761-18410/c

; Sequence 18410, Application US/09864761

; Patent No. US20020048763A1

GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-x-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1

SEQ ID NO 18410

LENGTH: 567

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AP000163.1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95

OTHER INFORMATION: SWISSPLOT HIT: P11308, EVALUE 9.00e-99

OTHER INFORMATION: EST_HUMAN HIT: R87572.1, EVALUE 0.00e+00

; OTHER INFORMATION: NT HIT: M17254.1, EVALUE 0.00e+00

US-09-864-761-18410

Query Match 11.4%; Score 199.4; DB 9; Length 567;

Best Local Similarity 79.5%; Pred. No. 1.7e-42;

Matches 236; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 546 AGGCAGCGGCAGATCCAGTTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGCGGAA 605

DB 523 AGGCAGTGGCCAGATCCAGCTTTGGCAGTTTCTCTCTGGAGCTCTCTTCGGACAGCTCCAA 464

QY 606 CGCCGCTCATCGCTGGAGGGCGGCCACCGCAGTTCAAGCTCACCGACCCCGACGA 665

DB 463 CTCACGCTCATCACCTGGGAAGGCCAACCGGGAGTTCAAGATACGGATCCCCGACGA 404

QY 666 GGTGGCGGACGCTGGGGCGAGCGCAAGAGCAAGCCCAATATGAACTACGACAAAGCTAAG 725

DB 403 GGTGGCCCGCGCTGGGGAGAGCGGAAGACAAACCAACATGAACATACGATAAGCTCAG 344

QY 726 TCGAGCACTGCGCTACTACTACGACAAACAAACATCATGACGACAGGTGCACGGCAAGCGCTA 785

DB 343 CGCGCCCTTCCGTTACTACTATGACAAAGAACATCATGACCAAGGTCCATGGGAAGCGCTA 284

QY 786 CGCCTACCGCTTTGACTTCCAGGGCTGGCAGAGGCTTCCGACGACCAACCCCGCGCA 842

DB 283 CGCCTACAGTTTCGACTTCCACGGGATCGCCAGGCCCTCCAGCCCCACCCCCCGGA 227

RESULT 9

US-09-864-761-20472/c

; Sequence 20472, Application US/09864761

; Patent No. US20020048763A1

GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-x-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

```
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20472
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000021.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EST HUMAN HIT: R87572.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUE 6.00e-88
; OTHER INFORMATION: NT HIT: M17254.1, EVALUE 0.00e+00
US-09-864-761-20472

Query Match 11.3%; Score 198.4; DB 9; Length 473;
Best Local Similarity 79.4%; Pred. No. 2.9e-42;
Matches 235; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 547 GCGAGGGGCGAGTCAGTTGGCAGATTTCTACTGGAGCTGCTGGCAGACCGCGGAAC 606
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 GCGAGTGGCCAGATCAGCTTTGGCAGTTTGGCAGTTCTCTGGAGCTCTGTGGACAGCTCCAAC 414
QY 607 GCCGCTGCATCGCTGGGAGGCGGCCACGCGAGTTCAAGCTCACCGACCCGCGAG 666
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
413 TCAGCTGCATCACCTGGAGAGCCACCAAGAGAGCAAAACCAACATGACGATCCGACGAG 354
QY 667 GTGGCGCGACGCTGGGCGGAGCGCAAGAGCAAGCCCAATATGAATCTACGACAGCTAAGT 726
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
353 GTGGCGCGCGCTGGGAGGCGGAAGAGCAAAACCAACATGACGATCCGACGCTCAGC 294
QY 727 CCAGCACTCGCTACTACTACACAAACATCATGAGCAAGTGGACGGGAGCGCTAC 786
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
293 CCGCGCTCTCGTTACTACTATGACAAAGAACATCATGACCAAGTCTCCATGGGAGCGCTAC 234
QY 787 GCCTACCGCTTTGACTTCCAGGCGCTGGCAGAGCTTGCAGCCACCCCGCGCA 842
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 GCCTACAGTTGACTTCCAGGGATCGCCAGGCTCCAGCCACCCCGCGCA 178

RESULT 10
US-10-029-386-25224/c
; Sequence 25224, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25224
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001731.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.5
```

```
; OTHER INFORMATION: NT HIT: g114780450, EVALUE 0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: R87572.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUE 3.00e-87
US-10-029-386-25224

Query Match 11.3%; Score 197.4; DB 12; Length 472;
Best Local Similarity 79.3%; Pred. No. 5.4e-42;
Matches 234; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 548 GCAGGGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGCGGAACG 607
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
472 GCAGTGGCCAGATCCAGCTTTGGCAGTTTCTCTGGAGCTCTGTGGAGCTCCAAC 413
QY 608 CCGGCTGCATCCGCTGGGAGGCGGCCACGCGAGTTCAAGCTCACCGACCCCGACGAGG 667
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
412 CCAGCTGCATCACCTGGGAGGCGGCCACCAAGCGGAGTTCAAGATGACGGATCCGACGAGG 353
QY 668 TGGCGCGACGCTGGGCGGAGCGCAAGAGCAAGCCCAATATGAATCTACGACAACTAAGTC 727
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 TGGCGCGCGCTGGGAGGCGGAAGAGCAAAACCAACATGAATACGATAAGCTCAGCC 293
QY 728 GAGCACTCGCTACTACTACGACAAACATCATGAGCAAGTGGACGCGCAAGCGCTACG 787
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 GCGCCCTCGTTACTACTATGACAAAGACATCATGACCAAGGTCCATGGGAAGCGCTACG 233
QY 788 CCTACCGCTTTGACTTCCAGGCGCTGGCAGAGCTTGCAGCCACCCCGCGCA 842
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 CCTACAAGTTGACTTCCAGGGATCGCCAGGCTCCAGCCCAACCCCGCGA 178

RESULT 11
US-10-007-926A-295
; Sequence 295, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 295
; LENGTH: 2957
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: friend leukemia virus integration 1 (FLI1)
; OTHER INFORMATION: gene.
US-10-007-926A-295

Query Match 11.1%; Score 194; DB 12; Length 2957;
Best Local Similarity 79.3%; Pred. No. 1e-40;
Matches 230; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 547 GCGAGGGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGCGGAAC 606
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1001 GGAAGCGGCGAGATCCAGCTGTGGCAATTCCTCTCGGAGCTGCTCTCCGACAGCGCAAC 1060
QY 607 GCCGCTGCATCGCTGGGAGGCGGCCACGCGAGTTCAAGCTCACCGACCCCGACGAG 666
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1061 GCCAGCTGTATCACTCTGGAGGGGAGCAACCGGGAGTTCAAAATGACGACCCCGATGAG 1120
QY 667 GTGGCGCGACGCTGGGCGGAGCGCAAGAGCAAGCCCAATATGAATACGACAACTAAGT 726
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1121 GTGGCCAGGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGAATTACGACAAGCTGAGC 1180
QY 727 CCAGCACTCGCTACTACTACACAAAAAACAATCATGCAAGCTGCAAGCGGAAAGCGCTAC 786
Db 1181 CGGGCCCTCCGGTTATTACTATGATATAAAACATTATGACCAAAAGTGCAGGCAAAAGATAT 1240
QY 787 GCGTACCGCTTTGACTTCCAGGCGCTGGGACACAGGCTTGCACGCCACCAACC 836
Db 1241 GCTTACAAATTGACTTCCAGGCGATTCGCCAGGCTCTGCAGCCACATCC 1290

RESULT 12

US-09-902-772-1
; Sequence 1, Application US/09902772
; Patent No. US20020164739A1
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: the Proteins
; CURRENT APPLICATION NUMBER: US/09/902,772
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US/08/878,177
; PRIOR FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA
US-09-902-772-1

Query Match 10.3%; Score 180; DB 10; Length 1447;
Best Local Similarity 74.0%; Pred. No. 3.7e-37;
Matches 228; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 529 AGCCCTGGCGGTACAGAAAGCGGAGCGGATCCAGTTGTGGCAGTTTCTACTGGAGCTG 588
Db 879 AGCCGCTCTTGCAAAATCCAGGGAGTGGGAGATACAGTATGCGAGTTCTACTGGAGCTT 938
QY 589 GTGGCAGACCGCGCAAGCCGGCTGCATCGGTGGGAGGGGCGCCACGGCAGTTCAAG 648
Db 939 CTGTGGGACAGCTCCAACTCCAACTGCATCACCCTGGGAGGGGCAAAATGGGGAGTTCAAG 998
QY 649 CTCACCGACCCCGACGAGTGGCGGACGCTGGGCGAGCGCAAGCAAGCCCAATATG 708
Db 999 ATGACAGACCTTGATGAAGTGGCTCGGCTGGGAGAGAGAAAGCAAACTTAACATG 1058
QY 709 AACTAGCAAGCTAAGTGCAGCACTGCGCTACTACTAGCAAAAAACATCATGAGCAAG 768
Db 1059 AACTATGACAAACTCAGCGCGTGCATTCGCTACTACTATGACAAAAATATTATGACTAAA 1118
QY 769 GTGCAGCGCAAGCGCTACCGCTTACCGCTTTGACTTCCAGGGCTGGCAGAGCTTGGCAG 828
Db 1119 GTTCATGTAACGCTATGCCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAG 1178
QY 829 CCACCACC 836
Db 1179 CCTCACCC 1186

RESULT 13

US-09-902-772-3
; Sequence 3, Application US/09902772
; Patent No. US20020164739A1
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: the Proteins
; CURRENT APPLICATION NUMBER: US/09/902,772
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US/08/878,177

; PRIOR FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: DNA
; ORGANISM: c-erg gene, chicken DNA
US-09-902-772-3

Query Match 10.3%; Score 180; DB 10; Length 1528;
Best Local Similarity 74.0%; Pred. No. 3.8e-37;
Matches 228; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 529 AGCCCTGGCGGTACAGAAAGCGGAGCGGATCCAGTTGTGGCAGTTTCTACTGGAGCTG 588
Db 960 AGCCGCTCTTGCAAAATCCAGGGAGTGGGAGATACAGTATGCGAGTTCTACTGGAGCTT 1019
QY 589 GTGGCAGACCGCGCAAGCCGGCTGCATCGGTGGGAGGGGCGCCACGGCAGTTTCAAG 648
Db 1020 CTGTGGGACAGCTCCAACTCCAACTGCATCACCCTGGGAGGGGCAAAATGGGGAGTTCAAG 1079
QY 649 CTCACCGACCCCGACGAGTGGCGGACGCTGGGCGGAGCGCAAGCAAGCCCAATATG 708
Db 1080 ATGACAGACCTTGATGAAGTGGCTCGGCTGGGAGAGAGAAAGCAAACTTAACATG 1139
QY 709 AACTAGCAAGCTAAGTGCAGCACTGCGCTACTACTAGCAAAAAACATCATGAGCAAG 768
Db 1140 AACTATGACAAACTCAGCGCGTGCATTCGCTACTACTATGACAAAAATATTATGACTAAA 1199
QY 769 GTGCAGCGCAAGCGCTACCGCTTACCGCTTTGACTTCCAGGGCTGGCAGAGCTTGGCAG 828
Db 1200 GTTCATGTAACGCTATGCCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAG 1259
QY 829 CCACCACC 836
Db 1260 CCTCACCC 1267

RESULT 14

US-10-210-120-41
; Sequence 41, Application US/10210120
; Publication No. US20030175736A1
; GENERAL INFORMATION:
; APPLICANT: Chinnaiyan, Arul M.
; APPLICANT: Rubin, Mark A.
; APPLICANT: Sreekumar, Arun
; TITLE OF INVENTION: Expression Profile of Prostate Cancer
; FILE REFERENCE: UM-07221
; CURRENT APPLICATION NUMBER: US/10/210,120
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US 60/309,581
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: US 60/334,468
; PRIOR FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-210-120-41

Query Match 7.9%; Score 138.8; DB 12; Length 441;
Best Local Similarity 69.6%; Pred. No. 1.8e-26;
Matches 188; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 544 AAAGCGCAGCGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTCTGGCAGACCGCGG 603
Db 103 ACAGAAAGTGACCTATTTCACCTGTGGCAGTTTCTCTCTGGAGCTCTATCAGCAAAATCC 162
QY 604 AACCGCGGCTGCATCGCTGGGAGGGGCGCCACGGCGAGTTCAAGCTCACCGACCCCGAC 663
Db 163 TGCCAGTCATTTCATCAGCTGGAGCTGGAGCGGAGTTTAAGCTCGC-CACCCCGAT 222

```
Qy 664 GAGTGGCGGACGCTGGGGGAGCGCAAGAGCAAGCCCAATATGAATACGACAAAGCTA 723
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
223 GAGTGGCGGCGCGGTGGGAAAGAGAGGAAATTAAGCCCAAGATGAATACGAGAAAGCTG 282
Qy 724 AGTCGAGCACTGCGCTACTACTACGACAAAACATCATGAGCAAGGTGCACGGCAAGCGC 783
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 AGCGGGGCTTACGCTACTATTACGACAAAGAACATCATCCAAAGACGTCGGGGAAGCGC 342
Qy 784 TACGCCTACCGCTTTGACTTCCAGGGCCTG 813
Db ||||| ||||| ||||| ||||| |||||
343 TACGTGTACCGCTTCGTGTGCGACCTCCAG 372
```

RESULT 15

```
US-09-918-995-27904
; Sequence 27904, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 27904
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(473)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27904
```

```
Query Match 7.9%; Score 138.8; DB 11; Length 473;
Best Local Similarity 69.6%; Pred. No. 1.8e-26;
Matches 188; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 544 AAGGCGAGCGGAGATCCAGTTGTCAGTTTCTACTGGAGCTGCTGGGAGACGGCGCG 603
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 ACAGGAAGTGGACCTATTACGCTGTGGCAGTTTCTCTGGAGCTGCTATCAGACAAATCC 224
Qy 604 AACCGCGGCTGCATCGCTGGGAGGCGGCCAGCGAGTTCAAGCTCACCGACCCCGAC 663
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 TGCCAGTCATTTCATCAGCTGGACTGGAGACGGATGGGAGTTTAAGCTCGCCGACCCCGAT 284
Qy 664 GAGTGGCGGACGCTGGGCGGAGCGCAAGAGCAAGCCCAATATGAATACGACAAAGCTA 723
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
285 GAGTGGCGGCGCGGTGGGAAAGAGGAAATTAAGCCCAAGATGAATACGAGAAAGCTG 344
Qy 724 AGTCGAGCACTGCGCTACTACTACGACAAAACATCATGAGCAAGGTGCACGGCAAGCGC 783
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
345 AGCGGGGCTTACGCTACTATTACGACAAAGAACATCATCCAAAGACGTCGGGGAAGCGC 404
Qy 784 TACGCCTACCGCTTTGACTTCCAGGGCCTG 813
Db ||||| ||||| ||||| ||||| |||||
405 TACGTGTACCGCTTCGTGTGCGACCTCCAG 434
```

Search completed: November 26, 2003, 16:38:30
Job time : 581 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 13:20:22 ; Search time 114 Seconds
(without alignments)
6783.362 Million cell updates/sec

Title: US-10-027-859-1

Perfect score: 1752
Sequence: 1 ttgttttaacaaacatgtt.....aaaaaaaaaaaaaaaaaaaaa 1752

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1752	100.0	1752	3	US-09-360-779-1
2	1752	100.0	1752	4	US-09-435-335-1
3	233.4	13.3	1754	4	US-09-604-978-1
4	233.4	13.3	1754	4	US-09-604-728-1
5	223	12.7	2532	4	US-09-620-312D-869
6	194	11.1	2938	2	US-08-343-443B-3
7	180	10.3	1447	3	US-08-878-177-1
8	180	10.3	1528	3	US-08-878-177-3
9	138.8	7.9	2268	3	US-09-344-579-1
10	134.8	7.7	2667	2	US-08-469-412A-1
11	134.8	7.7	2667	3	US-09-021-715-1
12	121	6.9	1734	3	US-09-146-969-5
13	121	6.9	1782	4	US-09-604-978-2
14	121	6.9	1782	4	US-09-604-728-2
15	115.4	6.6	1604	1	US-08-306-691B-43
16	115.4	6.6	1604	5	PCT-US93-06251-9
17	114.8	6.6	2064	3	US-08-875-944B-1
18	114.8	6.6	2064	3	US-09-116-049-3
19	114.8	6.6	2064	4	US-09-602-868A-1
20	114.8	6.6	2064	4	US-09-884-363-3
21	110	6.3	2265	2	US-09-213-767-1
22	108.8	6.2	2410	2	US-08-780-835B-1
23	108.8	6.2	2410	3	US-09-303-268-1
24	108.8	6.2	2410	3	US-09-116-049-1
25	108.8	6.2	2410	3	US-09-884-363-1
26	103.4	5.9	2544	2	US-08-469-412A-6
27	103.4	5.9	2544	3	US-09-021-715-6

28 97 5.5 7218 1 US-08-232-463-14 Sequence 14, Appli
29 90.4 5.2 1933 4 US-09-920-759-3 Sequence 3, Appli
30 90.4 5.2 1976 4 US-09-920-759-10 Sequence 10, Appli
31 86 4.9 665 4 US-09-920-759-11 Sequence 11, Appli
32 74 4.2 15788 4 US-09-920-759-13 Sequence 13, Appli
33 68 3.9 1894 4 US-09-570-593-1 Sequence 1, Appli
34 68 3.9 1905 3 US-09-055-113-2 Sequence 2, Appli
35 68 3.9 3317 4 US-09-570-593-12 Sequence 12, Appli
36 66.6 3.8 328 2 US-08-343-443B-5 Sequence 5, Appli
37 66.6 3.8 1907 4 US-09-300-958A-27 Sequence 27, Appli
38 66.6 3.8 1907 4 US-09-570-593-4 Sequence 4, Appli
39 65.8 3.8 65042 4 US-09-784-316-3 Sequence 3, Appli
40 65 3.7 1920 1 US-08-746-789A-1 Sequence 1, Appli
41 62.6 3.6 1364 1 US-08-306-691B-50 Sequence 50, Appli
42 62.6 3.6 1364 5 PCT-US93-06251-65 Sequence 65, Appli
43 60.6 3.5 320 3 US-09-165-264-14 Sequence 14, Appli
44 60.4 3.4 502 4 US-09-389-681-282 Sequence 282, App
45 60.4 3.4 502 4 US-09-620-405B-282 Sequence 282, App

ALIGNMENTS

RESULT 1

US-09-360-779-1

; Sequence 1, Application US/09360779

; Patent No. 6268216

; GENERAL INFORMATION:

; APPLICANT: Denaris, Evan S.

; APPLICANT: Pyodoro, Dmitry V.

; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds

; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases

; FILE REFERENCE: CASE-03828

; CURRENT APPLICATION NUMBER: US/09/360,779

; CURRENT FILING DATE: 1999-07-26

; EARLIER APPLICATION NUMBER: 60/094,264

; EARLIER FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1752

; TYPE: DNA

; ORGANISM: Rattus norvegicus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (112)..(1131)

US-09-360-779-1

Query Match 100.0%; Score 1752; DB 3; Length 1752;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGTTTTTAAACAATGTTTATTAGAAAAGTAAATAATTTGCATAGTCTTAGTACTTG 60
Db 1 TTGTTTTTAAACAATGTTTATTAGAAAAGTAAATAATTTGCATAGTCTTAGTACTTG 60
Qy 61 AACATCAAGTGTTTTCATGAACCGTGAGTATCTTCATGTAACAGTTCATAGTGAAGAC 120
Db 61 AACATCAAGTGTTTTCATGAACCGTGAGTATCTTCATGTAACAGTTCATAGTGAAGAC 120
Qy 121 CCAGTGGCGGCTCTCTGCGGGGAGAGGGTTCCAGCCCCCCCCCATCCCATCC 180
Db 121 CCAGTGGCGGCTCTCTGCGGGGAGAGGGTTCCAGCCCCCCCCCATCCCATCC 180
Qy 181 CTACACACTACTCTCTCAGTACACCGGCACCGGGATGGGCTGGGATGCAGCTCAGGAC 240
Db 181 CTACACACTACTCTCTCAGTACACCGGCACCGGGATGGGCTGGGATGCAGCTCAGGAC 240
Qy 241 CCCCTCCTCTCTCACCACACCCCTGGCTCCCGCTCCCGCAGGCTTTGCCGACCCGCG 300
Db 241 CCCCTCCTCTCTCACCACACCCCTGGCTCCCGCTCCCGCAGGCTTTGCCGACCCGCG 300

```
QY 301 GGGTCTACTTCTCCCTGTCAACCCAGTCAACACGGGGGGTATCGGCACCCCAAGCGCA 360
Db 301 GGGTCTACTTCTCCCTGTCAACCCAGTCAACACGGGGGGTATCGGCACCCCAAGCGCA 360
QY 361 AAGCTGAGTGGCCCGCCCGTGGGGTCCCGCCCATCTCCACCGCCAGTCCCGCGCAGCG 420
Db 361 AAGCTGAGTGGCCCGCCCGTGGGGTCCCGCCCATCTCCACCGCCAGTCCCGCGCAGCG 420
QY 421 ATGAGACAGAGCGGCACTCCAGCCCTGTGTGATCAACATGTACTACAGATCCCGTC 480
Db 421 ATGAGACAGAGCGGCACTCCAGCCCTGTGTGATCAACATGTACTACAGATCCCGTC 480
QY 481 GGAGATGCTCTTTTAAAGGAGGAGAGCCGAGCTGGGGCCGCTGAGCCCTCGGTA 540
Db 481 GGAGATGCTCTTTTAAAGGAGGAGAGCCGAGCTGGGGCCGCTGAGCCCTCGGTA 540
QY 541 CAGAAAGCGAGCGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGTGGCAGACCGC 600
Db 541 CAGAAAGCGAGCGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGTGGCAGACCGC 600
QY 601 CGGAACGGCGGCTGATCGGTGGAGGGCGGCCACGGGAGTTCAAGCTCACCGACCCC 660
Db 601 CGGAACGGCGGCTGATCGGTGGAGGGCGGCCACGGGAGTTCAAGCTCACCGACCCC 660
QY 661 GACGAGGTGGCGGCGAGCTGGGGCGAGCGCAAGAGCAAGCCCAATATGAATACGACAAG 720
Db 661 GACGAGGTGGCGGCGAGCTGGGGCGAGCGCAAGAGCAAGCCCAATATGAATACGACAAG 720
QY 721 CTAAGTCGAGCACTCGCTACTACTACGACAAAAACATCATGAGCAAGGTGCACGGCAAG 780
Db 721 CTAAGTCGAGCACTCGCTACTACTACGACAAAAACATCATGAGCAAGGTGCACGGCAAG 780
QY 781 CGTAGCGCTACCGCTTTGACTTCAGGGCGCTGGCAGAGCTTGCAGCCACACCGCG 840
Db 781 CGTAGCGCTACCGCTTTGACTTCAGGGCGCTGGCAGAGCTTGCAGCCACACCGCG 840
QY 841 CACGCCACGCGCGCTGCGCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCACTT 900
Db 841 CACGCCACGCGCGCTGCGCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCACTT 900
QY 901 TACAAGCTCCCGGCTGGTCTGCTGCTCACTGCGCTTCCCGGCGCTTCCAAACTCAACCTT 960
Db 901 TACAAGCTCCCGGCTGGTCTGCTGCTCACTGCGCTTCCCGGCGCTTCCAAACTCAACCTT 960
QY 961 ATGGCAGCTCGCGCGGCTGGCGCGCGCTGCTTCTTACTGCGCTGCTCCCAAGCGC 1020
Db 961 ATGGCAGCTCGCGCGGCTGGCGCGCGCTGCTTCTTACTGCGCTGCTCCCAAGCGC 1020
QY 1021 ACCGCGCTGCGCGCGCGCGCTGCTGCTTACCCAAACCGCGGCTTGCAGCGCCCTCC 1080
Db 1021 ACCGCGCTGCGCGCGCGCGCGCTGCTTACCCAAACCGCGGCTTGCAGCGCCCTCC 1080
QY 1081 GGGCCCTTTGGCGGCTGGCGCGCGCTTCCGACTTGGGGGGTCAATATCACTAGACGGGA 1140
Db 1081 GGGCCCTTTGGCGGCTGGCGCGCGCTTCCGACTTGGGGGGTCAATATCACTAGACGGGA 1140
QY 1141 CGGCGGCTGAGTGGGGCTCTCCACACAGCGAGTACCAATCCCATCTCATCTGG 1200
Db 1141 CGGCGGCTGAGTGGGGCTCTCCACACAGCGAGTACCAATCCCATCTCATCTGG 1200
QY 1201 GAGGAGCCCGAAGATTTCCCGAGCTTCTTTACACAGATTTCTGTTGCAGCAGCGCT 1260
Db 1201 GAGGAGCCCGAAGATTTCCCGAGCTTCTTTACACAGATTTCTGTTGCAGCAGCGCT 1260
QY 1261 CCCAGCCAGGAGAAAGATGGGAAGCTCTGAGGCTTCTTGAATAAGAGCTTCC 1320
Db 1261 CCCAGCCAGGAGAAAGATGGGAAGCTCTGAGGCTTCTTGAATAAGAGCTTCC 1320
QY 1321 AGGCTCCCATATATCATCAACCCAGGAGGCTGCTGCCACTTTAAATTTTCTT 1380
Db 1321 AGGCTCCCATATATCATCAACCCAGGAGGCTGCTGCCACTTTAAATTTTCTT 1380
QY 1381 CCAAGTCTCCAGATTTCTGGAACTCCCGCTTTTTTTTTTCTTCTTCTCACCTGGAGCCCTGC 1440
```

```
Db 1381 CCAAGTCTCCAGATTTCTGGAACTCCCGCTTTTTTTTTTCTTCTCACCTGGAGCCCTGC 1440
QY 1441 CTTCTCTTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTTTTTCTCTCTCTCTCTC 1500
Db 1441 CTTCTCTTTATGACCCCTAGTTTCTGTTTGTGTTTTTTTTTTTTTCTCTCTCTCTC 1500
QY 1501 ATTTTCTTCTCCACGACTACTCCAAAGCGTAGTACCTCGGTAGTACCTCGAGGCT 1560
Db 1501 ATTTTCTTCTCCACGACTACTCCAAAGCGTAGTACCTCGGTAGTACCTCGAGGCT 1560
QY 1561 TCTCACACTCCCTTTTCGGGATAGAAAGCATCAAAAACATCTCTGCTGTTGTCATC 1620
Db 1561 TCTCACACTCCCTTTTCGGGATAGAAAGCATCAAAAACATCTCTGCTGTTGTCATC 1620
QY 1621 CCTATCCCAACACTCTGCTTCCGCTTCCATACCACTCTCTGGCCCAAGGACCTCG 1680
Db 1621 CCTATCCCAACACTCTGCTTCCGCTTCCATACCACTCTCTGGCCCAAGGACCTCG 1680
QY 1681 TCTGTATATATTCCTTTTCAGCCCATTAAGATCCAAAGTTCAAAAAATAAAAAA 1740
Db 1681 TCTGTATATATTCCTTTTCAGCCCATTAAGATCCAAAGTTCAAAAAATAAAAAA 1740
QY 1741 AAAAAAATAA 1752
Db 1741 AAAAAAATAA 1752

RESULT 2
US-09-435-335-1
; Sequence 1, Application US/09435335
; Patent No. 6384204
; GENERAL INFORMATION:
; APPLICANT: Deneris, Evan S.
; APPLICANT: Fyodorov, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; FILE REFERENCE: CASE-04027
; CURRENT APPLICATION NUMBER: US/09/435.335
; CURRENT FILING DATE: 1999-11-05
; EARLIER APPLICATION NUMBER: 09/360.779
; EARLIER FILING DATE: 1999-07-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(1131)
; US-09-435-335-1

Query Match 100.0%; Score 1752; DB 4; Length 1752;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTGTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGCAATAGTCTTAGTACTTG 60
Db 1 TTTGTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGCAATAGTCTTAGTACTTG 60
QY 61 AACATCAAGTGTATTCATGAACCGTAGATATCTTCATGATAAACAGTCTTAGATGGAAGAC 120
Db 61 AACATCAAGTGTATTCATGAACCGTAGATATCTTCATGATAAACAGTCTTAGATGGAAGAC 120
QY 121 CCAGGTGGCGCTCTCTGGGGAGAGGTTCCAGCCCCCAGCCCCCTCAGCCCCATCCC 180
Db 121 CCAGGTGGCGCTCTCTGGGGAGAGGTTCCAGCCCCCAGCCCCCTCAGCCCCATCCC 180
QY 181 CTCACAGCTCACTCTCTCAGTACACCGGCAACCGGATGGGCTGGGATGCAGTCCAGGAC 240
Db 181 CTCACAGCTCACTCTCTCAGTACACCGGCAACCGGATGGGCTGGGATGCAGTCCAGGAC 240
```


QY	241	CCCTCCCTCCTCACCACACCTTGGCTGGCCGCTCCCGCAGGCGCTTGC CGGACCCCGCGC	300
DB	241	CCCTCCCTCCTCACCACACCTTGGCTGGCCGCTCCCGCAGGCGCTTGC CGGACCCCGCGC	300
QY	301	CGCTACTCTTCCCTGTCACCACAGTCAACAAGGCGGTATCGGCACCCCAAGCGCA	360
DB	301	CGCTACTCTTCCCTGTCACCACAGTCAACAAGGCGGTATCGGCACCCCAAGCGCA	360
QY	361	AAGCTGACGTGCCCGCCCGTGGCTGCCCGCCATCTCCACCGCCCACTCCCGCGCAGC	420
DB	361	AAGCTGACGTGCCCGCCCGTGGCTGCCCGCCATCTCCACCGCCCACTCCCGCGCAGC	420
QY	421	ATGAGACAGGGACCTCCAGCCCTGCTGATCAACATGTACTACAGATCCCGTC	480
DB	421	ATGAGACAGGGACCTCCAGCCCTGCTGATCAACATGTACTACAGATCCCGTC	480
QY	481	GGAGATGTCCTTTTAAGGAAGGAAGACCGAGCTGGGGCGCTGAGCCCTCGCGTA	540
DB	481	GGAGATGTCCTTTTAAGGAAGGAAGACCGAGCTGGGGCGCTGAGCCCTCGCGTA	540
QY	541	CAGAAAGCAGCGGCAGATCCAGTTGTGGAGTTTCTACTGGAGCTGCTGGCAGCGC	600
DB	541	CAGAAAGCAGCGGCAGATCCAGTTGTGGAGTTTCTACTGGAGCTGCTGGCAGCGC	600
QY	601	CGGAACCGCGCTGCATCGCTGGAGGGCGGCCACAGCGAGTTCAAGCTCACCAGCCC	660
DB	601	CGGAACCGCGCTGCATCGCTGGAGGGCGGCCACAGCGAGTTCAAGCTCACCAGCCC	660
QY	661	GACGAGTGGCGCAGCGCTGGGCGAGCGCAAGAGCAAGCCCAATATGAATACGACAAG	720
DB	661	GACGAGTGGCGCAGCGCTGGGCGAGCGCAAGAGCAAGCCCAATATGAATACGACAAG	720
QY	721	CTAAGTCGAGCACTCGCTACTACTACGACAAAAACATCATGACAGGTGCACGCAAG	780
DB	721	CTAAGTCGAGCACTCGCTACTACTACGACAAAAACATCATGACAGGTGCACGCAAG	780
QY	781	CGCTACGGCTACCGTTTGACTTCAGGGCCTGGCACAAGGTTGGCAGCACACCGCGC	840
DB	781	CGCTACGGCTACCGTTTGACTTCAGGGCCTGGCACAAGGTTGGCAGCACACCGCGC	840
QY	841	CACGCCACCGCGCGTGGCGCGCAGCGCAGCGCGCGCCAGGATGGCGCACTT	900
DB	841	CACGCCACCGCGCGTGGCGCGCAGCGCAGCGCGCGCCAGGATGGCGCACTT	900
QY	901	TACAAGCTCCGGTGTGTGGCTCCACTGCGCTTCCCGCGCTCTCCAACTCAACCTT	960
DB	901	TACAAGCTCCGGTGTGTGGCTCCACTGCGCTTCCCGCGCTCTCCAACTCAACCTT	960
QY	961	ATGCAAGCTCGCGGGGTGGCGCGCTGGCTTCTTACTGGCTGTGCCACGCC	1020
DB	961	ATGCAAGCTCGCGGGGTGGCGCGCTGGCTTCTTACTGGCTGTGCCACGCC	1020
QY	1021	ACCGCGCTCGCGCGCACCGCTGCGCTTACCCCAACCCCGGCTTGCAGCCCCCTCC	1080
DB	1021	ACCGCGCTCGCGCGCACCGCTGCGCTTACCCCAACCCCGGCTTGCAGCCCCCTCC	1080
QY	1081	GGGCGCTTTGGCGGGTGGCGCGCTTCGCACTTGGGGGTCAATATCACTAGACGGGA	1140
DB	1081	GGGCGCTTTGGCGGGTGGCGCGCTTCGCACTTGGGGGTCAATATCACTAGACGGGA	1140
QY	1141	CGGCGGGTGAATGGGGCTCTCCCAACAGCGAGTGA CCAATCCCATCTCATCTCG	1200
DB	1141	CGGCGGGTGAATGGGGCTCTCCCAACAGCGAGTGA CCAATCCCATCTCATCTCG	1200
QY	1201	GAGGAGCCCGAAGATTCCCGAGCTTCTTTACACAGATTTCGTGAGCGCGCT	1260
DB	1201	GAGGAGCCCGAAGATTTCGCGAGCTTCTTTACACAGATTTCGTGAGCGCGCT	1260
QY	1261	CCGAGCCCGGGAAGAGGATGGAGCGCTCTGAGCTTCTCTTGAATACGAGGCTTC	1320
DB	1261	CCGAGCCCGGGAAGAGGATGGAGCGCTCTGAGCTTCTCTTGAATACGAGGCTTC	1320

RESULT 3

US-09-604-978-1/c

; Sequence 1, Application US/09604978

; Patent No. 6455674

GENERAL INFORMATION:

APPLICANT: Einat. Paz

Skaliter. Rami

TITLE OF INVENTION: HYPOXIA-REGULATED GENES

NUMBER OF SEQUENCES: 11

NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:

ADDRESSEE: KOHN & ASSOCIATES

STREET: 30500 No. 6455674thwestern Hwy. Suite 401

CITY: Farmington Hills

STATE: Michigan

COUNTRY: U.S.

ZIP: 48334

COMPUTER READABLE FORM:

```

;
;
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0. Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/604,978

FILING DATE: 28-Jun-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Kohn, Kenneth I.

REGISTRATION NUMBER: 30.955

REGISTRATION NUMBER: 5013
REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (248) 539-5050

TELEPHONE: (248) 5395
TELEFAX: (248) 5395

/ INFORMATION FOR SEO ID NO: 1:
1EEBFAA: (Z48) 33930

: SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:
LENGTH: 1754 base pairs

LENGTH: 1/34 base pairs
TYPE: nucleic acid

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-604-978-1
Query Match 13.3%; Score 233.4; DB 4; Length 1754;
Best Local Similarity 99.6%; Pred. No. 1.3e-45;
Matches 234; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGTCATAGGCTTTAGTACTTG 60
Db 1733 TTGTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGTCATAGGCTTTAGTACTTG 1674

Qy 61 AACATCAAGTGTATTTCATGAACCGTGCAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 120
Db 1673 AACATCAAGTGTATTTCATGAACCGTGCAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 1614

Qy 121 CCAGGTGGCGCTCCTCTGCGGGAGAGGGTTCCAGCCCCCCCCCAGCCCCCTCAGCCCCCATCCC 180
Db 1613 CCAGGTGGCGCTCCTCTGCGGGAGAGGGTTCCAGCCCCCCCCCAGCCCCCTCAGCCCCCATCCC 1554

Qy 181 CTCACAGCTCACTCCTCCAGTACACCGGCACCGGGATGGCTGGGATGCAGCTCC 235
Db 1553 CTCACAGTTCACCTCCTCCAGTACACCGGCACCGGGATGGCTGGGATGCAGCTCC 1499

RESULT 4
US-09-604-728-1/c
; Sequence 1, Application US/09604728
; Patent No. 6555667
; GENERAL INFORMATION:
; APPLICANT: Einat, Paz
; TITLE OF INVENTION: HPOXIA-REGULATED GENES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 No. 6555667thwestern Hwy., Suite 401
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/604,728
; FILING DATE: 28-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,112
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 0168.00034
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 5395055
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-604-728-1
Query Match 13.3%; Score 233.4; DB 4; Length 1754;
Best Local Similarity 99.6%; Pred. No. 1.3e-45;
Matches 234; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGTCATAGGCTTTAGTACTTG 60
Db 1733 TTGTTTAAACAACATGTTTATTAGAAAAGTAAAAATATTGTCATAGGCTTTAGTACTTG 1674

Qy 61 AACATCAAGTGTATTTCATGAACCGTGCAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 120
Db 1673 AACATCAAGTGTATTTCATGAACCGTGCAGTATCTTTCATGTAAACAGTTCTAGATGGAAGAC 1614

Qy 121 CCAGGTGGCGCTCCTCTGCGGGAGAGGGTTCCAGCCCCCCCCCAGCCCCCTCAGCCCCCATCCC 180
Db 1613 CCAGGTGGCGCTCCTCTGCGGGAGAGGGTTCCAGCCCCCCCCCAGCCCCCTCAGCCCCCATCCC 1554

Qy 181 CTCACAGCTCACTCCTCCAGTACACCGGCACCGGGATGGCTGGGATGCAGCTCC 235
Db 1553 CTCACAGTTCACCTCCTCCAGTACACCGGCACCGGGATGGCTGGGATGCAGCTCC 1499

RESULT 5
US-09-620-312D-869
; Sequence 869, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: dt_FL_genes Version 1.0
; SEQ ID NO 869
; LENGTH: 2532
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2532)
; US-09-620-312D-869

Query Match 12.7%; Score 223; DB 4; Length 2532;
Best Local Similarity 72.2%; Pred. No. 4.3e-43;
Matches 366; Conservative 0; Mismatches 70; Indels 71; Gaps 3;

Qy 546 AGCGACGGCGCAGATCCAGATTGTCGAGTTTCTATGAGAGTGTGCGCAGACCGCGCGAA 605
Db 2096 ATGCAGCGGACAGATCCAGCTGTGCGAGTTTCTGTGAGAGTGTGCTGGCTGACCGCGCGAA 2155
```


Db 939 CTGTCGACAGCTCCAACTCCAATGCATCACCTGGGAGGCGCAAAATGGGAGTTCAAG 998
 QY 649 CTCACGACCCGACGAGGTGGCGGACGCTGGGGCGAGCGCAAGCAAGCCCAATATG 708
 Db 999 ATCAGACGCTCATGAAGTGGTTCGGCGTGGGAGAGAGGAAAGCAACCTAATCATG 1058
 QY 709 AACTAGCACAAGCTAAGTCGAGCACTGCGCTACTACTACGACAAACAAACATCATGACGACG 768
 Db 1059 AACTATGACAACTCAGCGTGCACTTCGCTACTACTATGACAAAATATTATGACTAAA 1118
 QY 769 GTCCACGGCAAGCGCTACGCTACCGCTTTGACTTCCAGGGCCTGGCAGAGGCTTGCCAG 828
 Db 1119 GTTCATGGTAAACGCTATGCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAG 1178
 QY 829 CCACCAACC 836
 Db 1179 CCTCACCC 1186

RESULT 8
 US-08-878-177-3
 ; Sequence 3, Application US/08878177
 ; Patent No. 6294354
 ; GENERAL INFORMATION:
 ; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
 ; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
 ; FILE OF INVENTION: the Proteins
 ; FILE REFERENCE: Chugai Seiyaku Kabushiki Kaisha 5001
 ; CURRENT APPLICATION NUMBER: US/08/878,177
 ; CURRENT FILING DATE: 1997-06-18
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 1528
 ; TYPE: DNA
 ; ORGANISM: c-erg gene, chicken DNA
 ; US-08-878-177-3

Query Match 10.3%; Score 180; DB 3; Length 1528;
 Best Local Similarity 74.0%; Pred. No. 4.6e-33;
 Matches 228; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 529 AGCCCTCGGTACAGAAAGCGACGCGGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTG 588
 Db 960 AGCCGTCTTGCAATCCAGGAGTGGGAGATACAGCTATGCGAGTTCTACTGGAGCTT 1019
 QY 589 CTGGCAGACCGCGCAACCGCGCTGCATCGGTGGGAGGCGGCGCGAGTTTCAAG 648
 Db 1020 CTGTCGGACAGCTCCAACTCCAATCGATCACCTGGGAGGCGCACAAATGGGAGTTTCAAG 1079
 QY 649 CTCACCGACCCGACGAGGTGGCGGACGCTGGGGCGAGCGCAAGCAAGCCCAATATG 708
 Db 1080 ATGACAGACCTTGATGAAGTGGTTCGGGTGGGAGAGAGGAAAGCAAACTAATATG 1139
 QY 709 AACTAGCACAAGCTAAGTCGAGCACTGCGCTACTACTACGACAAACAAACATCATGACGACG 768
 Db 1140 AACTATGACAACTCAGCGTGCACTTCGCTACTACTATGACAAAATATTATGACTAAA 1199
 QY 769 GTCCACGGCAAGCGCTACGCTTACCGCTTTGACTTCCAGGGCCTGGCAGAGGCTTGCCAG 828
 Db 1200 GTTCATGGTAAACGCTATGCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAG 1259
 QY 829 CCACCAACC 836
 Db 1260 CCTCACCC 1267

RESULT 9
 US-09-344-579-1
 ; Sequence 1, Application US/09344579
 ; Patent No. 6054316
 ; GENERAL INFORMATION:
 ; APPLICANT: Brenda F. Baker

; APPLICANT: Lex M. Cowbert
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF ETS-2 EXPRESSION
 ; FILE REFERENCE: RTS-0063
 ; CURRENT APPLICATION NUMBER: US/09/344,579
 ; CURRENT FILING DATE: 1999-06-25
 ; NUMBER OF SEQ ID NOS: 47
 ; SEQ ID NO 1
 ; LENGTH: 2268
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (291)..(1700)
 ; US-09-344-579-1

Query Match 7.9%; Score 138.8; DB 3; Length 2268;
 Best Local Similarity 69.6%; Pred. No. 2.8e-23;
 Matches 188; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 544 AAAGCGAGCGGCGAGATCCAGTTGGGAGTTTCTACTGGAGCTGTGGCAGACCGCGCG 603
 Db 1362 ACAGGAAGTGGACCTATTTCAGCTGTGGCAGTTTCTCTGGAGCTGTATCAGACAAATCC 1421
 QY 604 AACGCCGCTGCATCGCTGGGAGGCGCGCCAGCGAGTTCAAGCTCACCGACCGCGAC 663
 Db 1422 TGCCAGTCAATTCATCAGCTGGACTGGAGACGGATGGAGCTCGCCGACCCCGAT 1481
 QY 664 GAGGTGGCGCGACGCTGGGGCGAGCGCAAGAGCAAGCCCAATATGAATCAACAAAGCTA 723
 Db 1482 GAGGTGGCGCGCGCTGGGGAAAGAGGAAATAAGCCCAAGTGAATACGAGAAGCTG 1541
 QY 724 AGTCGACACTGCGCTACTACTACGACAAACAAACATCATGAGCAAGTGCACGCGCAAGCGC 783
 Db 1542 AGCCGGGCGTTACGCTACTATTACGACAAAGAAACATCATCCACAAGACGTCGGGGAAGCGC 1601
 QY 784 TACGCTACCGCTTTGACTTCCAGGCGCTG 813
 Db 1602 TACGTGTACCGCTTCTGTGCGACCTCCAG 1631

RESULT 10
 US-08-469-412A-1
 ; Sequence 1, Application US/08469412A
 ; Patent No. 5856125
 ; GENERAL INFORMATION:
 ; APPLICANT: Mavrothalassitis, George J.
 ; APPLICANT: Blair, Donald G.
 ; APPLICANT: Fisher, Robert J.
 ; APPLICANT: Beal Jr., Gregory J.
 ; APPLICANT: Athanasios, Metropi A.
 ; APPLICANT: Sgouras, Dionysios N.
 ; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/469,412A
 ; FILING DATE: 05-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Garrett-Wackowski, Eugenia
 ; REGISTRATION NUMBER: 37,330
 ; REFERENCE/DOCKET NUMBER: 015280-229000

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2667 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..1769
; OTHER INFORMATION: /note= "human ERF (ETS2 Repressor
; Factor) cDNA"
; US-08-469-412A-1
;
; Query Match 7.7%; Score 134.8; DB 2; Length 2667;
; Best Local Similarity 68.2%; Pred. No. 2.7e-22;
; Matches 187; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
;
; Qy 547 GGCAGCGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTCTGGCAGACCGCGAAC 606
; Db |||||
; Qy 607 GCGCGTGCATCCGCTGGAGGGCGCCAGCGGAGTTCAAGCTACCGACCCCGACGAG 666
; Db |||||
; Qy 667 GTGCGCGAGCTGGGCGGAGCGCAAGACCAATATCAATACGACAAAGCTAAGT 726
; Db |||||
; Qy 727 CGAGCACTGCGCTACTACTACGACAAAAACATCATGACGAAGGTGACGCAAGCGCTAC 786
; Db |||||
; Qy 787 GCTACCGCTTGTGACTTCAGGGCCCTGGCAGAG 820
; Db |||||
; Qy 820 ACCTACAAGTTCAATTTCAACAACTGGTGTGG 462
;
; RESULT 11
; US-09-021-715-1
; Sequence 1, Application US/09021715
; Patent No. 6194547
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; Blair, Donald G.
; Fisher, Robert J.
; Beal Jr., Gregory J.
; Athanasios, Meropi A.
; Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/021,715
; APPLICATION NUMBER: US/09/021,715
; FILING DATE: 10-Feb-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2667 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..1769
; OTHER INFORMATION: /note= "human ERF (ETS2 Repressor
; Factor) cDNA"
; US-09-021-715-1
;
; Query Match 7.7%; Score 134.8; DB 3; Length 2667;
; Best Local Similarity 68.2%; Pred. No. 2.7e-22;
; Matches 187; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
;
; Qy 547 GGCAGCGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTCTGGCAGACCGCGAAC 606
; Db |||||
; Qy 607 GCGCGTGCATCCGCTGGAGGGCGCCAGCGGAGTTCAAGCTACCGACCCCGACGAG 666
; Db |||||
; Qy 667 GTGCGCGAGCTGGGCGGAGCGCAAGACCAATATCAATACGACAAAGCTAAGT 726
; Db |||||
; Qy 727 CGAGCACTGCGCTACTACTACGACAAAAACATCATGACGAAGGTGACGCAAGCGCTAC 786
; Db |||||
; Qy 787 GCTACCGCTTGTGACTTCAGGGCCCTGGCAGAG 820
; Db |||||
; Qy 820 ACCTACAAGTTCAATTTCAACAACTGGTGTGG 462
;
; RESULT 12
; US-09-146-969-5/C
; Sequence 5, Application US/09146969
; Patent No. 6228585
; GENERAL INFORMATION:
; APPLICANT: Dieckgraefe, Brian K.
; TITLE OF INVENTION: Gene Markers for Chronic Mucosal Injury
; FILE REFERENCE: 04255.75314
; CURRENT APPLICATION NUMBER: US/09/146,969
; CURRENT FILING DATE: 1998-09-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-146-969-5
;
; Query Match 6.9%; Score 121; DB 3; Length 1734;
; Best Local Similarity 83.4%; Pred. No. 3.9e-19;
; Matches 176; Conservative 0; Mismatches 25; Indels 10; Gaps 3;
;
; Qy 1 TTGTGTTTAAACAAACATGTTTATTAGAAAAGT-AAAAATATTGCATAGCTCTTAGTACTT 59
; Db |||||
; Qy 1725 TTGTGTTTAAACAAACATGTTTATTAGAAAAGTAAAAAATATTGCATAGCTCTTAGTACTT 166
; Db |||||
; Qy 60 GAACATCAAGTGTATTTCATGAACCGGTGAG-----TATCTTCTGTAACAGTTCTTAGATG 114
; Db |||||

```

Db 1665 GAACATCAAGTGATTTATCATGAACAGTGAGTATCTTATCTTATGTAAGTAAACAGTTCTAGATG 1606
 QY 115 GAAGACCCAGTGGCTCTCTGCGGGAGAGGTTCCAGCCCCCAGCCCTCAGCCC 174
 Db 1605 GAAGACCCAGATGGCACTCTCTCCCGGGAGGGGTTCCAGCCCCCAGCCCTCTC-----AGCC 1550
 QY 175 CATCCCTCACAGCTCACTCTCTCCAGTACAC 205
 Db 1549 CTTCCCTGCCAGCTCAACTCTGCAGTACAC 1519

RESULT 13
 US-09-604-978-2/c
 ; Sequence 2, Application US/09604978
 ; Patent No. 6455674
 ; GENERAL INFORMATION:
 ; APPLICANT: Einat, Paz
 ; Skaliter, Rami
 ; TITLE OF INVENTION: HYPOXIA-REGULATED GENES
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: KOHN & ASSOCIATES
 ; STREET: 30500 No. 6455674thwestern Hwy., Suite 401
 ; CITY: Farmington Hills
 ; STATE: Michigan
 ; COUNTRY: U.S.
 ; ZIP: 48334

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/604,978
 FILING DATE: 28-Jun-2000
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/138,112
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Kohn, Kenneth I.
 REGISTRATION NUMBER: 30,955
 REFERENCE/DOCKET NUMBER: 0168.00034
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (248) 539-5050
 TELEFAX: (248) 5395055
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1782 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-604-978-2
 Query Match 6.9%; Score 121; DB 4; Length 1782;
 Best Local Similarity 83.4%; Pred. No. 4e-19;
 Matches 176; Conservative 0; Mismatches 25; Indels 10; Gaps 3;

QY 1 TTGTTTTAACAAACATGTTTATTAGAAAAGT-AAAAATATTGCATAGGTCTTAGTACTT 59
 Db 1767 TTGTTTTAACAAACATGTTTATTAGAAAAGTAAAAAATATTGCATAGGTCTTAGTACTT 1708
 QY 60 GAACATCAAGTGATTTATCATGAACCGTGAG-----TATCTTCATGTAACAGTTCTAGATG 114
 Db 1707 GAACATCAAGTGATTTATCATGAACAGTGAGTATCTTATCTTCATGTAACAGTTCTAGATG 1648
 QY 115 GAAGACCCAGTGGCGCTCTCTGCGGGAGAGGTTCCAGCCCCCAGCCCTCAGCCC 174
 Db 1647 GAAGACCCAGATGGCACTCTCTCCCGGGAGGGGTTCCAGCCCCCAGCCCTCTC-----AGCC 1592

QY 175 CATCCCTCACAGCTCACTCTCTCCAGTACAC 205
 Db 1591 CTTCCCTGCCAGCTCAACTCTGCAGTACAC 1561
 RESULT 14
 US-09-604-728-2/c
 ; Sequence 2, Application US/09604728
 ; Patent No. 6555667
 ; GENERAL INFORMATION:
 ; APPLICANT: Einat, Paz
 ; Skaliter, Rami
 ; TITLE OF INVENTION: HYPOXIA-REGULATED GENES
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: KOHN & ASSOCIATES
 ; STREET: 30500 No. 6555667thwestern Hwy., Suite 401
 ; CITY: Farmington Hills
 ; STATE: Michigan
 ; COUNTRY: U.S.
 ; ZIP: 48334

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/604,728
 FILING DATE: 28-Jun-2000
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/138,112
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Kohn, Kenneth I.
 REGISTRATION NUMBER: 30,955
 REFERENCE/DOCKET NUMBER: 0168.00034
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (248) 539-5050
 TELEFAX: (248) 5395055
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1782 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-604-728-2
 Query Match 6.9%; Score 121; DB 4; Length 1782;
 Best Local Similarity 83.4%; Pred. No. 4e-19;
 Matches 176; Conservative 0; Mismatches 25; Indels 10; Gaps 3;

QY 1 TTGTTTTAACAAACATGTTTATTAGAAAAGT-AAAAATATTGCATAGGTCTTAGTACTT 59
 Db 1767 TTGTTTTAACAAACATGTTTATTAGAAAAGTAAAAAATATTGCATAGGTCTTAGTACTT 1708
 QY 60 GAACATCAAGTGATTTATCATGAACCGTGAG-----TATCTTCATGTAACAGTTCTAGATG 114
 Db 1707 GAACATCAAGTGATTTATCATGAACAGTGAGTATCTTATCTTCATGTAACAGTTCTAGATG 1648
 QY 115 GAAGACCCAGTGGCGCTCTCTGCGGGAGAGGTTCCAGCCCCCAGCCCTCAGCCC 174
 Db 1647 GAAGACCCAGATGGCACTCTCTCCCGGGAGGGGTTCCAGCCCCCAGCCCTCTC-----AGCC 1592
 QY 175 CATCCCTCACAGCTCACTCTCTCCAGTACAC 205
 Db 1591 CTTCCCTGCCAGCTCAACTCTGCAGTACAC 1561

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 13:17:02 ; Search time 3862 Seconds
(without alignments)
11025.749 Million cell updates/sec

Title: US-10-027-859-1

Perfect score: 1752
Sequence: 1 ttgttttaacaacatgtt.....aaaaaaaaaaaaaaaaaaaa 1752

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_estum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hcc.*

9: gb_estl.*

10: gb_est2.*

11: gb_hcc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_man.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vri.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	472.4	27.0	599	12	BI438572 iC24a08.x
2	452.2	25.8	590	14	CA867675 iR31f10.y
c 3	415.8	23.7	495	14	CA771025 iO69g10.x
4	415.4	23.7	584	14	CA771327 iO69g10.y

c 5	414.4	23.7	472	14	CA867392	CA867392 iR31f10.x
c 6	404	23.1	436	9	AW488372	AW488372 UI-M-BH3-
c 7	373.2	21.3	516	12	BI194956	BI194956 602943780
c 8	361.6	20.6	531	10	BE948739	BE948739 UI-M-BH3-
c 9	358.2	20.4	379	10	BE948133	BE948133 UI-M-BH3-
c 10	356.4	20.3	632	10	BE262829	BE262829 601151586
11	354	20.2	1120	13	BU603585	BU603585 AGENCOURT
c 12	334.8	19.1	649	10	BE312349	BE312349 601148529
c 13	333.6	19.0	451	14	CB069127	CB069127 iR31c07.x
14	329	18.8	582	14	CB069363	CB069363 iR31c07.y
15	318	18.2	458	10	BF470469	BF470469 UI-M-BH3-
16	288.6	16.5	624	13	BQ777436	BQ777436 iL48b07.y
c 17	283.4	16.2	635	28	AQ937962	AQ937962 NB6-699R
c 18	271.2	15.5	480	13	BQ788110	BQ788110 iL48b07.x
c 19	270	15.4	321	12	BM128563	BM128563 iF12h07.y
c 20	260.8	14.9	430	12	BI963862	BI963862 ie65f04.x
c 21	255.8	14.6	675	12	BM440515	BM440515 pgrln.pk0
c 22	251.2	14.3	377	28	AQ939613	AQ939613 NR3-088R
c 23	244.2	13.9	671	13	BU417183	BU417183 603671376
c 24	242.6	13.8	289	12	BM128343	BM128343 iF12h07.x
c 25	233.4	13.3	322	9	AI577542	AI577542 UI-R-Y0-v
c 26	233.4	13.3	675	13	BQ199014	BQ199014 UI-R-EA0-
c 27	233.4	13.3	700	13	BQ209348	BQ209348 UI-R-DY1-
c 28	233.4	13.3	759	14	CA509882	CA509882 UI-R-F50-
c 29	231.8	13.2	346	9	AI555662	AI555662 UI-R-C2p-
c 30	231.8	13.2	499	9	AI144692	AI144692 UI-R-BT0-
c 31	231.8	13.2	656	14	CD371354	CD371354 UI-R-DY0-
c 32	231.8	13.2	673	12	BM384149	BM384149 UI-R-D20-
c 33	230.2	13.1	290	9	AI555747	AI555747 UI-R-C2p-
c 34	225.4	12.9	358	9	AI235276	AI235276 EST331838
c 35	225.4	12.9	456	9	AI408716	AI408716 EST337007
c 36	220.8	12.6	694	13	BQ210597	BQ210597 UI-R-DY1-
c 37	219.8	12.5	634	28	AQ939612	AQ939612 NR3-088C
c 38	216.8	12.4	251	9	AI105175	AI105175 EST214464
c 39	205	11.7	524	9	AI794504	AI794504 fc45b11.y
c 40	203	11.6	1866	11	AK050922	AK050922 Mus muscu
c 41	203	11.6	3064	11	AK078113	AK078113 Mus muscu
c 42	198.8	11.3	466	13	BQ270970	BQ270970 iK11f04.x
c 43	198.2	11.3	602	14	CB582161	CB582161 AMGNNUC:N
c 44	198.2	11.3	616	14	CB579500	CB579500 AMGNNUC:N
c 45	197.8	11.3	709	14	CB269538	CB269538 1008445 H

ALIGNMENTS

RESULT 1	BI438572/c	BI438572	iC24a08.x3	HR85	islet Homo sapiens	cdna 3', similar to TR:070132	599 bp	linear	EST 21-AUG-2001
LOCUS	BI438572	O70132	ETS	DOMAIN	TRANSCRIPTION FACTOR PET-1. ;	mrna	EST 21-AUG-2001		
DEFINITION	BI438572	GI:15263262							
ACCESSION	BI438572								
VERSION	EST.								
KEYWORDS	EST.								
SOURCE	EST.								
ORGANISM	Homo sapiens (human)								
REFERENCE	Homo sapiens								
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 599)								
	Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.								
TITLE	Endocrine Pancreas Consortium								
JOURNAL	Unpublished								
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue								
	Endocrine Pancreas Consortium								
	Harvard University, Howard Hughes Medical Institute								
	Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138								
	Tel: 617-495-1812								

Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@im.wustl.edu)
 High quality sequence stop: 475.
 Location/Qualifiers
 1..599
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="Purified pancreatic islet"
 /lab_host="DH10B"
 /clone_lib="HR85 islet"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
 NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
 Size selected on agarose gel. Average insert size -1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,
 Metabolism Div. (Alan Permutt Lab), Washington University
 School of Medicine, Box 8127, 660 South Euclid Ave., St.
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
 314-362-1916, Fax: 314-747-2692."
 82 a 191 c 223 g 100 t 3 others

FEATURES

source

Query Match 27.0%; Score 472.4; DB 12; Length 599;
 Best Local Similarity 88.8%; Pred. No. 3e-39;
 Matches 532; Conservative 0; Mismatches 64; Indels 3; Gaps 2;
 QY 289 CCGGACCCGGCGCTTACTTCTCCCTGTACCCACAGTCACACGGCGGGTATCGGC 348
 DB 599 CCGGACCCGGCGCGTTTCTTCTCTGTACCCCGGTGCNCGGGGGGATCGGT 540
 QY 349 ACC-CAAAGCGCAAAGTGAAGTGTGCCCCCGT--GCGTCCCCCCCCATCTCCACCGCC 405
 DB 539 GCCACGAGCGAAAGCGTGTGCGCCCCCTTCCCGGTGNCCTCCATCTCCACCGCC 480
 QY 406 CAGTCCCGGACGATGAGACAGAGCGGACCTCCGACCCCTCTCATCATGTATC 465
 DB 479 CAGTCCCGGCGGATGAGACAGAGCGGCGCTCCCGACCCCTCTCATCAATGTATC 420
 QY 466 TTACCATGATCCCGTGGAGATGGTCTTTTAAAGAGGGAAGAGCCCGAGCTGGGGCGG 525
 DB 419 CTGCCAGATCCGCTGGAGACGGTCTTCAAGAGCGGGAAGAACCCGAGCTGGGGCGG 360
 QY 526 CTGAGCCCTGGGTCAGAAAGGCGAGCGGAGATCCAGTTGTGGCAGTTTCTACTGGAG 585
 DB 359 CTGAGCCCGCGGTTTCAAGAAAGGCGAGCGAGATCCAGTGTGGCAGTTTCTGTGAG 300
 QY 586 CTGCTGGCAGACCGCGAAACCGCGTGCATCGGTGGGAGGGCGGCCACGGCGAGTTC 645
 DB 299 CTGCTGGTGTACCGCGGAAACCGCGGTGCATCGGTGGGAGGGCGGTTCACGGCGAGTTC 240
 QY 646 AAGCTCACCGACCCCGACGAGTGGCGGAGCGCTGGGCGGAGCGCAAGAGCAAGCCCAAT 705
 DB 239 AAGCTCACCGACCCCGACGAGTGGCGGCGGTGGGCGGAGCGCAAGAGCAAGCCCAAC 180
 QY 706 ATGAACCTACGACAGCTAAGTCGAGCATGCGCTACTACTACGACAAACATCATGAGC 765
 DB 179 ATGAACCTACGACAGCTAAGTCGAGCATGCGCTACTACTACGACAAACATCATGAGC 120
 QY 766 AAGTGGACGCAAGCGCTACGCTTACCGCTTTGACTTCCAGGGCTTGGCAGAGCTTGC 825
 DB 119 AAGTGGATGCAAGCGCTACGCTTACCGCTTTGACTTCCAGGGCTTGGCAGAGCTTGC 60
 QY 826 CAGCCACCAACCGGCGACGCCCAAGCGCCCGCTGCGCGCGCGGAGCGGACGCGCGCC 884
 DB 59 CAGCGCGCGCGCGGACGCTCATGCGCGCGCGCGAGTGTGCGCGCGCGCGCGCGCG 1

RESULT 2

CA867675
 LOCUS
 DEFINITION

ir3if10.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6546788 5'
 similar to TR:070132 070132 ETS DOMAIN TRANSCRIPTION FACTOR PET-1.
 ; mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS

CA867675.1 GI:27319224
 EST.
 CA867675.1

ORGANISM
 SOURCE

Homo sapiens (human)
 Homo sapiens

REFERENCE
 AUTHORS

Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
 Lemishka,L., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
 Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,J., Blisstein,A.,
 Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
 M., Gibbons M., McCann,R., Cole,R., Tsagataishvili,R., Williams,T.,
 Jackson,Y. and Bowers,Y.
 Endocrine Pancreas Consortium
 Unpublished
 Other ESTs: ir3if10.x1

JOURNAL
 COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@im.wustl.edu)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 455.
 Location/Qualifiers
 1..599
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6546788"
 /tissue_type="Purified pancreatic islet"
 /lab_host="DH10B"
 /clone_lib="HR85 islet"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
 NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
 Size selected on agarose gel. Average insert size -1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,
 Metabolism Div. (Alan Permutt Lab), Washington University
 School of Medicine, Box 8127, 660 South Euclid Ave., St.
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
 314-362-1916, Fax: 314-747-2692."
 103 a 224 c 179 g 84 t

BASE COUNT
 ORIGIN

Query Match 25.8%; Score 452.2; DB 14; Length 590;
 Best Local Similarity 89.5%; Pred. No. 3.4e-37;
 Matches 509; Conservative 0; Mismatches 58; Indels 2; Gaps 2;

QY 240 CCCCCTCCCTCTCACACACCTTGGCTGCCGCTCCCGCAGGCTTCCGCGACCCCGC 299
 DB 22 CCCCCTCAACCCCGCACCGCCGGGTGCCCTCCCGCAGGCTTCCGCGACCCCGC 81
 QY 300 GGCGCTACTCTTCCCTGTGTACCCACAGTCACACGGCGGGTATCGGCACC-CAAACGC 358
 DB 82 GGCGCTTCTCTCTGTGTACCCCGGTGCCTTCCGCGGGGATCGGTGCACCGACGC 141
 QY 359 CAAAG-CTGAGTGTCCCCCGGTGCGGTGCCCTCCATCTCCACCGCCAGTCCCCGCA 417
 DB 142 CAAAGCTGTCTCGCCCTTCCCGTCCCGCCCATCTCCACCGCCAGTCCCCGCG 201

352	TTCAAGACGGGAAGAACCCGAGCTGGGGCCGTGAGCCCGCGGTTCAAGAAAGGCAGC	293
553	GGGCAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGCTGGCAGACCGCGAAACCGCGGC	612
292	GGACAGATCCAGCTGTGGCAGTTTCTCTGGAGCTGCTGGCTGACCGCGAAACCGCGGC	233
613	TGCATCCGCTGGAGGGCGCGCACGGCGAGTTTCAAGCTCACCGACCCCGACGAGGTGGCG	672
232	TGCATCCGCTGGAGGGCGGTCACGGCGAGTTTCAAGCTCACGGACCCCGACGAGGTGGCG	173
673	CGAGCTGGGGCGAGCGCAAGAGCAAGCCCAATATGAACTACGACACAGCTTAAGTTCGACGA	732
172	CGCGGTGGGGCGAGCGCAAGAGCAAGCCCAATGAACACGACAAAGCTGAGCGCGGCC	113
733	CTCGCTTACTACTACGACAAAAACATCATGAGCAAGGTGACGCGCAAGCGCTACGCGCTAC	792
112	CTCGCTTACTACTACGACAAACATCATGAGCAAGGTGTCATGGCAAGCGCTACGCGCTAC	53
793	CGCTTTGACTTTCAGGGCGCTGGACACAGGCTTGCCAGCGCACACCCCGCGCAGC	844
52	CGCTTCGACTTCAGGGCGCTGGCGAGGCGCTTGCAGCGCGCGCGCGCGCGCAGC	1

RESULT 6					
AW488372/c					
LOCUS	AW488372	436 bp	mRNA	linear	EST 24-FEB-2000
DEFINITION	UI-M-BH3-arp-g-06-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone UI-M-BH3-arp-g-06-0-UI 3', mRNA sequence.				

FEATURES
source

BASE COUNT	63 a	142 c	141 g	90 t
ORIGIN				
Query Match		23.1%	Score 404;	DB 9; Length 436;
Best Local Similarity		95.4%	Pred. No. 3.4e-32;	

RESULT 7
BI194956
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS


```

BASE COUNT      TAG_SEQ=None found"
ORIGIN          85 a 172 c 170 g 103 t 1 others

Query Match      20.6%; Score 361.6; DB 10; Length 531;
Best Local Similarity 94.8%; Pred. No. 5.6e-28;
Matches 384; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

QY 459 CATGTACTACAGATCCCGTGGAGATGGTCTTTTAAAGGAAGGAAGAGCCGAGCTG 518
DB 405 CTTNTGCCGCGAGATCCCGTGGAGATGGTCTTTTAAAGGAAGGAAGAGCCGAGCTG 346
QY 519 GGGCCGCTGAGCCCTGCGGTACAGAAAGGAGCGGGAGATCCAGTTGTGCGAGTTTCT 578
DB 345 GGGCCGCTGAGCCCGCGGTACAGAAAGGAGCGGGAGATCCAGTTGTGCGAGTTTCT 286
QY 579 ACTGAGCTGTGGCAGACCGCGCAACCGCGCTGCATCGTGGAGGGCGGCCAGCG 638
DB 285 CTTGAGCTGTGCGAGATCGCGCAACCGCGTGTGATCGGTGGAGGGCGGCCAGCG 226
QY 639 CGAGTTCAAGCTACCGACCCCGACGAGGTGGCGACGCTGGGGCGAGCGCAAGAGCAA 698
DB 225 CGAGTTCAAGCTACCGACCCCGACGAGGTGGCGCGCTGGGGCGAGCGCAAGAGCAA 166
QY 699 GCCAATATGACTACGACAGCTTAAGTCGAGCACTGCGCTACTACTACGACAAAAACAT 758
DB 165 GCCTAATGACTACGACAGCTTAAGTCGCGCGCTGCGCTACTACTACGACAAAAACAT 106
QY 759 CATGAGCAAGGTGCACGCGCAAGCGCTACCGCTTACCGCTTTGACTT-CCAGGCGCTGGCAC 817
DB 105 CATGAGCAAGGTGCACGCGCAAGCGCTACCGCTTACCGCTTTGACTTCCAGGCGCTGGCGC 46
QY 818 AGGCTTGCCAGCCACCAACCGCGCAAGCGCTGGCGCGCTGGCGCGCTGGCGCG 862
DB 45 AGGCTTGCCAGCCACCAACCGCGCAAGCGCTGGCGCGCTGGCGCGCG 1

RESULT 9
BE948133/c
LOCUS
DEFINITION
  UI-M-BH3-awn-c-08-0-UI-s1 NIH BMAP M S4 Mus musculus cDNA clone
  UI-M-BH3-awn-c-08-0-UI 3', mRNA sequence.
ACCESSION
  BE948133.1 GI:10525892
VERSION
  EST.
KEYWORDS
  Mus musculus (house mouse)
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 379)
  Bonaldo, M.F., Lennon, G. and Soares, M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
JOURNAL
  97044477
MEDLINE
  8889548
PUBMED
  Contact: Chin, H
  National Institute of Mental Health
  6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
  20892-9643, USA
  Tel: 301 443 1706
  Fax: 301 443 9890
  Email: mEST@mail.nih.gov
  Oligo-dT track not found, Not 1 site shown in beginning of sequence
  is likely internal to the message. cDNA Library Preparation: M.B.
  Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
  clones from RESEARCH GENETICS. It should be noted that Bento Soares
  is generating a small number of additional specialized
  non-redundant arrays of BMAP cDNAs whose availability will be
  considered under appropriate and limited collaborative arrangements
  Seq primer: M13 Forward
  POLYA=No.
  Location/Qualifiers

source
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="UI-M-BH3-awn-c-08-0-UI"
  /dev_stage="27-32 days"
  /lab_host="DH10B (Life Technologies)"
  /clone_lib="NIH BMAP M S4"
  /note="Vector: p773D-Pac (Pharmacia) with a modified
  polylinker; Site 1: Not 1; Site 2: Eco RI; The
  NIH BMAP M S4 library is a subtracted library of a series,
  ultimately derived from a mixture of individually tagged
  normalized libraries from ten regions of the mouse brain
  (cerebellum, brain stems, olfactory bulb, hypothalamus,
  cortex, amygdala, basal ganglia, pineal gland, striatum,
  hippocampus) after a series of subtractions to reduce the
  representation of cDNAs from which ESTs had already been
  generated. The following serially subtracted libraries
  were generated in this process: NIH BMAP M S4,
  NIH BMAP M S3.3, NIH BMAP M S3.2, NIH BMAP M S3.1,
  NIH BMAP M S2, NIH BMAP M S1. The subtracted library
  (NIH BMAP M S4) was constructed as follows: PCR amplified
  cDNA inserts from NIH BMAP M S3.3, NIH BMAP M S3.2, and
  NIH BMAP M S3.1 clones from which 3' ESTs had been derived
  was used as a driver in a hybridization with a pool of
  the NIH BMAP M S3.3, NIH BMAP M S3.2, and NIH BMAP M S3.1
  libraries in the form of single-stranded circles. The
  remaining single-stranded circles (subtracted library)
  was purified by hydroxyapatite column chromatography,
  converted to double-stranded circles and electroporated
  into DH10B bacteria (Life Technologies) to generate the
  NIH BMAP M S4 library. This procedure has been previously
  described (Bonaldo, Lennon and Soares, Genome Research
  6:791-806, 1996)
  TAG_SEQ=None found"
  54 a 128 c 118 g 79 t

BASE COUNT      54 a 128 c 118 g 79 t
ORIGIN

Query Match      20.4%; Score 358.2; DB 10; Length 379;
Best Local Similarity 96.6%; Pred. No. 1.7e-27;
Matches 366; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 484 GATGCTCTTTTAAAGGAAGGAAGAGCCGAGCTGGGGCGCTGAGCCCTGCGGTACAG 543
DB 379 GATGCTCTTTTAAAGGAAGGAAGAGCCGAGCTGGGGCGCTGAGCCCTGCGGTACAG 320
QY 544 AAAGCGAGCGGCGAGATCCAGTTGTGCGAGTTTCTACTGGAGCTGCTGGCAGACCGCGCG 603
DB 319 AAAGCGAGCGGCGAGATCCAGTTGTGCGAGTTTCTCTCTGGAGCTGCTGGCAGATCGCGCG 260
QY 604 AACGCCGCTGCATCGCTGGAGGGCGGCACCGCGAGTTCAAGCTCACCAGCCCGGAC 663
DB 259 AACGCCGCTGCATCGCTGGAGGGCGGCACCGCGAGTTCAAGCTCACCAGCCCGGAC 200
QY 664 GAGTGGCGGCGAGCTGGCGGCGAGCGCAAGAGCCCAATATGAACTACGCAAGCTA 723
DB 199 GAGTGGCGGCGGCGCTGGGGCGGAGCGCAAGAGCCCTTAACATGAATACGCAAGCTA 140
QY 724 AGTCGAGCACTGCGCTACTACTACGCAAAAAACATATGAGCAAGGTCGACGGCAAGCGC 783
DB 139 AGTCGCGGCTGCGCTACTACTACGCAAAAAACATATGAGCAAGGTCGACGGCAAGCGC 80
QY 784 TAGCCTTACCGCTTTGACTTCCAGGGCTGGCAGAGCTTGCAGCCACCAACCCCGCGCAC 843
DB 79 TAGCCTTACCGCTTTGACTTCCAGGGCTGGCAGAGCTTGCAGAGGCTTGCAGCCACCAACCCCGCGCAC 20
QY 844 GCCACGCGCCCGCTGCGCG 862
DB 19 GCCACGCGCCCGCTGCGCGCG 1

RESULT 10

```

QY	872	CGGCAGC - CGCCGCCCCAGGATGGCGCACTTTACAAGCT - CCCGGCTGGTCTGGCTCCACT	929
Db	385	CGCCCGCTGGCGCGCCAGACGCGCGCTCTACAAGCTGCCCGCTGGCTCGCCCGCT	444
QY	930	GCCCTTCGCCGCTCTCAAACCTAACTTAAGGAGACCTCGGCGCGGTGGCGGCC-G	988
Db	445	GCCCTTCGCCGCTCTCAAACCTAACTTAAGGAGACCTCGGCGCGGTGGCGGCC	504
QY	989	CTGGCTTCTTACTTGCGCTGGTCCCAAGCCACCCTCGCTGCGCGGCCACC-----	1041
Db	505	CGGCTTCTCTACTTGCGCTGGTCCCAAGCCACCCTCGCTGCGCGGCCACC-----	1041
QY	1042	GCTGCGCTTACCCAACCCCGGCTTTCGACGCCCCCTCCCGGCGCTTTCGCGCGGTGCC	1101
Db	565	GCGCGCTTACCCCGCTTTCGACGCCCCCTCCCGGCGCTTTCGCGCGGTGCC	624
QY	1102	GCGCGTTC 1109	
Db	625	GCAGCCTC 632	
RESULT 11	BUE03585	BU603585	1120 bp mRNA linear EST 20-SEP-2002
LOCUS	AGENCOURT_8934291 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6499070	5', mRNA Sequence.	
DEFINITION	BU603585	BU603585.1 GI:23255344	
ACCESSION	BU603585	Homo sapiens (human)	
VERSION	BU603585	Homo sapiens	
KEYWORDS	EST.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
SOURCE		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
ORGANISM		NIH-MGC http://img.nci.nih.gov/	
REFERENCE		1 (bases 1 to 1120)	
AUTHORS		National Institutes of Health, Mammalian Gene Collection (MGC)	
TITLE		Unpublished	
JOURNAL		Contact: Robert Strausberg, Ph.D.	
COMMENT		Email: cgabps@mail.nih.gov	
		Tissue Procurement: NCI	
		CDNA Library Preparation: Michael Brownstein Laboratory	
		DNA Sequencing by: Agencourt Bioscience Corporation	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LNL at:	
		http://image.lnl.gov	
		Plate: LLCW2682 row: 1 column: 15	
		High quality sequence stop: 290.	
		Location/Qualifiers	
		1..1120	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone=IMAGE:6499070"	
		/tissue.type="mixed (pool of 40 RNAs)"	
		/lab_host="DH10B (TI-phase-resistant)"	
		/clone_lib="NIH MGC 142"	
		/note="Vector: pDNR-LIB; Site 1: SfiI (ggcattatggcc);	
		Site 2: SfiI (ggcgcttcgccc); Double-stranded cDNA was	
		prepared from a pool of 40 cell line polyA+ RNAs (bladder -	
		2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -	
		4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,	
		kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,	
		ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary	
		gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were	
		used in cloning as follows:	
		5'-AAGCATGTATCAACGAGTGGCCATTACGGCGGG-3' and	
		5'-ATTCATAGCGGCGGCGGCATG-dt(30)NN-3'. Full-length	
		enriched library was constructed using the Clontech	
		Creator SMART kit and size-selected to contain the >0.5 kb	
		size fraction (other fractions present in NIH MGC 141)	
		library created in the laboratory of Dr. Robert Strausberg,	
		NIH-MGC.	

BASE COUNT 174 a 397 c 353 g 193 t 3 others
 ORIGIN NIH). Note: this is a NIH MGC Library."

Query Match 20.2%; Score 354; DB 13; Length 1120;
 Best Local Similarity 87.6%; Pred. No. 1.7e-27;
 Matches 424; Conservative 0; Mismatches 50; Indels 10; Gaps 3;
 QY 428 AGAGCGGCGACCTCCAGCCCTGCTGATCAATGATACCTACAGATCCCGTGGAGATG 487
 Db 3 AGAGCGGCGCTCCAGCCCTGCTGATCAATGATACCTGCGAGATCCCGTGGAGAG 62
 QY 488 GTCTTTTAAAGAGGAAGAGCGGAGCTGGGGCGCTGAGCCCTGCGGTACAGAAAG 547
 Db 63 GTCTCTTCAAGGACGGAAGAACCCGAGCTGGGGCGCTGAGCCCGCGGTTCAGAAAG 122
 QY 548 GAGCGGGGAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGTGGCAGACCGCGGAACG 607
 Db 123 GAGCGGACAGATCCAGCTGTGGCAGTTTCTGCTGGAGCTGTGGCTGACCGCGGAACG 182
 QY 608 CCGGCTGTCATCGCTGGGAGGGCGCCACGGGAGTTCAAGCTCACGACCCCGAGAGG 667
 Db 183 CCGGCTGTCATCGCTGGGAGGGCGGTACGGGAGTTCAAGCTCACGACCCCGAGAGG 242
 QY 668 TGGCGGACGCTGGGCGGAGCGCAAGCAAGCCCAATATGAATACGACAAGCTTAAGTC 727
 Db 243 TGGCGGCGGTGGGCGGAGCGCAAGCAAGCCCAATGAATACGACAAGCTTAAGTC 302
 QY 728 GAGCACTGCGCTACTACTACGACAA-AAACATCATGAGCAAGTGCACGCAAGCGCTAC 786
 Db 303 GGGCCCTGCGCTACTACTACGACAA-AAACATCATGAGCAAGTGCACGCAAGCGCTAC 362
 QY 787 GCTACCGCTTTGACTTCCA-GGGCTGGCAGAGGTTGCCAGCCACCCCGCAGCG 845
 Db 363 GCTACCGCTTGCATTCAGGGGCTGGCGAGGCTGCCAGCGCGCGCGCACGT 422
 QY 846 CCAGCGCGGCTGGCGCGCGCAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 905
 Db 423 TCATGCCCCCG-----GCGCAGCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 474
 QY 906 GCTC 909
 Db 475 GGGC 478

RESULT 12
 BE312349 649 bp mRNA linear EST 26-OCT-2000
 LOCUS 601148529F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3163855 5',
 DEFINITION mRNA sequence.
 BE312349
 BE312349.1 GI:9130901
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 649)
 NIH-MGC http://imgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLM125 row: i column: 08
 High quality sequence stop: 467.
 Location/Qualifiers
 1. .649

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3163855"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_19"
 /note="Organ: brain; Vector: pOTB7; Site: 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5',
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 93 a 250 c 212 g 94 t
 ORIGIN
 Query Match 19.1%; Score 334.8; DB 10; Length 649;
 Best Local Similarity 78.2%; Pred. No. 2.4e-25;
 Matches 482; Conservative 0; Mismatches 117; Indels 17; Gaps 6;
 QY 546 AGCGAGCGGACAGATCCAGTTGTGGCAGTTTCTACTGGAGCTGTGGCTGACCGCGGAA 605
 Db 3 AGAAAGCGGACAGATCCAGCTGTGGAGTTTCTGCTGGAGCTGTGGCTGACCGCGGAA 62
 QY 606 CGCCGCTGTCATCGCTGGGAGGGCGCCACGGCGAGTTCAAGCTCACCGACCCCG-ACG 664
 Db 63 CGCGGCTGTCATCGCTGGGAGGGCGGTACCGCGAGTTCAAGCTCACCGACCCCGTAGC 122
 QY 665 AGGTGGCGGACGCTGGGGCGAGCGCAAGAGCAAGCCCAATATGAATACGACAAGCTAA 724
 Db 123 AGGTGGCGGCGGTGGGGCGAGCGCAAGAGCAAGCCCAATGAATACGACAAGCTGA 182
 QY 725 GTCGAGCACTGCGCTACTACTACGACAAACATCATGAGCAAGTGCATGCGCAAGCGCT 784
 Db 183 GCGCGCGCTTGGCTACTACTACGACAAAGAAATCATGAGCAAGTGCATGCGCAAGCGCT 242
 QY 785 AGCGCTACCGCTTTGACTTTCAGGCGCTTGGCAGAGGCTTGGCAGCACCACCCCGCGCACG 844
 Db 243 AGCGCTACCGCTTTCAGGCGCTTGGCAGAGGCTTGGCAGCACCACCCCGCGCGCA 302
 QY 845 CCCAGCGCGCGCTGCGCG-----CGCGCGAGCGGAGCGCGCGCGCGCGCGCGCGCG 896
 Db 303 CGCTCAATGCGCGCGCGCGAGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 362
 QY 897 ACTTTACAGCTCCCGGCTGGTCTCGCTCCACTGCGCTTCCCGGCGCTTCCGAACTCAA 956
 Db 363 TCTACGAAGCTGCGCGCGCGCGCTCGC-CCGGTGGCGCTTCCCGGCGCTTCCGAACTCAA 421
 QY 957 CCTTATGCGAGCTGCGCGCGCGCTGGCGCGCGCTGCGCTTCTTACTGCGCTGGTCCC-- 1014
 Db 422 CCTCATGCGCGCTGCGCGCGGTGCGCGCGCGCTTCTCTACTTGGCGCGCGCTG 481
 QY 1015 --AAGCCACCGCGCTGCGCGCGCGCGCGCGCGCTGCGCTCTTA-CCCAACCCCGGCGTTCAG 1071
 Db 482 GCG 541
 QY 1072 CCCCTTCCCGGCGCGCTTGGCGCGCGCGCGCGCGCGCTTCCACTTGGGGGCTCATTTATCAC 1131
 Db 542 CCCCCGCGCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 599
 QY 1132 TAGACGGGACGCGCG 1147
 Db 600 TAGACGGGCGCGCG 615
 RESULT 13
 CB069127/c 451 bp mRNA linear EST 21-JAN-2003
 LOCUS is13c07.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6364788 3',
 DEFINITION Similar to TR:Q99581 Q99581 FEV PROTEIN.; mRNA sequence.
 ACCESSION CB069127


```

VERSION      CB069127.1  GI:27813647
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 451)
AUTHORS      Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
              Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
              Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,J., Blustain,A.,
              Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
              M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
              Jackson,Y. and Bowers,Y.
TITLE        Endocrine Pancreas Consortium
JOURNAL      Unpublished
COMMENT      Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
              Endocrine Pancreas Consortium
              Harvard University, Howard Hughes Medical Institute
              Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
              MA 02138
              Tel: 617-495-1812
              Fax: 617-495-8557
              Email: dmelton@biohp.harvard.edu
              Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
              Washington University Genome Sequencing Center For information on
              obtaining a clone please contact: Dr. Hiroshi Inoue
              (hinoue@im.wustl.edu)
              Seq primer: -40RP from Gibco
              High quality sequence stop: 439.
FEATURES     Location/Qualifiers
              1..451
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6364788"
                /tissue_type="Purified pancreatic islet"
                /lab_host="DH10B"
                /clone_lib="HR85 islet"
                /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
              NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
              Size-selected on agarose gel. Average insert size ~1kb. 5'
              xhoI site was destroyed after directional cloning.
              Amplified once. Contact information: Hiroshi Inoue, MD,
              Metabolism Div. (Alan Permutt Lab), Washington University
              School of Medicine, Box 8127, 660 South Euclid Ave., St.
              Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
              314-362-1916, Fax: 314-747-2692."
BASE COUNT   74 a 147 c 147 g 83 t
ORIGIN
Query Match      19.0%; Score 333.6; DB 14; Length 451;
Best Local Similarity 91.2%; Pred. No. 4.4e-25;
Matches 354; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 461 TGTTACTACAGATCCGTCGGAGATGGTCTTTTAAAGGAAGCGAAGAGCCCGAGCTGGG 520
    |||
Db 389 TGTGCTTGCAGATCCGTCGGAGAGCGTCTTCAAGAGCGGAAGACCCGAGCTGGG 330
    |||
QY 521 GGCCTGTGAGCCCTGCGGTACAGAAAGGAGCGGGGAGATCCAGTTGTGCGCAGTTTCTAC 580
    |||
Db 329 GGCCTGTGAGCCCGCGGTTCAGAAAGGAGCGGACAGATCCAGTGTGCGAGTTTCTGC 270
    |||
QY 581 TGGAGCTGTGGCAGACCGCGCAACCGCGGTGCATCCGTGGGAGGGCGGCACGGCG 640
    |||
Db 269 TGGAGCTGTGGCTGACCGCGCAACCGCGGTGCATCCGTGGGAGGGCGGTTCACGGCG 210
    |||
QY 641 AGTTCAAGCTCACCGACCCCGCAGAGGTGGCGGAGCTGGGGCGAGCGCAAGAGCAAGC 700
    |||
Db 209 AGTTCAAGCTCACCGACCCCGCAGAGGTGGCGGCGGTGGGCGAGCGCAAGAGCAAGC 150
    |||
QY 701 CCATATGACTACGACAGCTTAAGTCGAGCACTGCGCTACTACTACGACAAACATCA 760
    |||
Db 149 CCAACATGAACACGACGCTAGCGCGGCCCTGCGCTACTACTACGACAAACATCA 90
    |||

```

```

QY 761 TGAGCAAGGTGCACGCAAGCGCTACGCTACCGCTTTCAGCTTCCAGGGCCTCGCACAGG 820
    |||
Db 89 TGAGCAAGGTGCATGCAAGCGCTACGCTACCGCTTTCAGCTTCCAGGGCCTCGGCAGG 30
    |||
QY 821 CTTGCCAGCCACACCCCGCGCACGCCCA 848
    |||
Db 29 CTTGCCAGCCCGCGCGCGCGCGCGCTCA 2
    |||
RESULT 14
LOCUS    CB069363
DEFINITION
CB069363
ACCESSION
CB069363
VERSION
CB069363.1  GI:27813883
KEYWORDS
EST.
SOURCE   Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 582)
AUTHORS      Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
              Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
              Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,J., Blustain,A.,
              Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
              M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
              Jackson,Y. and Bowers,Y.
TITLE        Endocrine Pancreas Consortium
JOURNAL      Unpublished
COMMENT      Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
              Endocrine Pancreas Consortium
              Harvard University, Howard Hughes Medical Institute
              Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
              MA 02138
              Tel: 617-495-1812
              Fax: 617-495-8557
              Email: dmelton@biohp.harvard.edu
              Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
              Washington University Genome Sequencing Center For information on
              obtaining a clone please contact: Dr. Hiroshi Inoue
              (hinoue@im.wustl.edu)
              Seq primer: -40UP from Gibco
              High quality sequence stop: 477.
FEATURES     Location/Qualifiers
              1..582
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6364788"
                /tissue_type="Purified pancreatic islet"
                /lab_host="DH10B"
                /clone_lib="HR85 islet"
                /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
              NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
              Size-selected on agarose gel. Average insert size ~1kb. 5'
              xhoI site was destroyed after directional cloning.
              Amplified once. Contact information: Hiroshi Inoue, MD,
              Metabolism Div. (Alan Permutt Lab), Washington University
              School of Medicine, Box 8127, 660 South Euclid Ave., St.
              Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
              314-362-1916, Fax: 314-747-2692."
BASE COUNT   98 a 195 c 205 g 84 t
ORIGIN
Query Match      18.8%; Score 329; DB 14; Length 582;
Best Local Similarity 90.9%; Pred. No. 1e-24;
Matches 350; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 461 TGTTACTACAGATCCGTCGGAGATGGTCTTTTAAAGGAAGCGAAGAGCCCGAGCTGGG 520
    |||

```

```

Db      197  TGCTGCTTGCAGATCCCGTGGAGACGGTCTTCTTCAAGGACGGGAAGAACCCGAGCTGGG 256
QY      521  GGCCGCTGAGCCCTGGGTACAGAAAGGACGGCGGAGATCCAGTTGTGGCAGTTTCTAC 580
Db      257  GGCCGCTGAGCCCGCGGTTCAGAAAGGACGGCGAGATCCAGCTGTGGCAGTTTCTGC 316
QY      581  TGGAGCTGCTGGCAGACCGCGGAAACCGCGGTTCATCGCTGGAGGGCGGCCACGGCG 640
Db      317  TGGAGCTGCTGGGTGACCGCGGAAACCGCGGTTCATCGCTGGAGGGCGGTTCACGGCG 376
QY      641  AGTTCAAGCTCACGACCGGACGGTGGCGGAGCTGGCGGAGCGCCGAGAGCAAGC 700
Db      377  AGTTCAAGCTCACGACCGGACGGTGGCGGAGCTGGCGGAGCGCCGAGAGCAAGC 436
QY      701  CCAATATGAACACGACGAGCTAAGTCGAGCACTGCGCTACTACTACGACAAACATCA 760
Db      437  CCAACATGAACACGACGAGCTAGCGCGCGCTGCTACTACTACGACAAACATCA 496
QY      761  TGAGCAAGGTGACGCGGACGCTAGCGCTTTCAGCTTTCAGCGGCTGGCAGAGG 820
Db      497  TGAGCAAGGTGACGCGGACGCTAGCGCTTTCAGCTTTCAGCGGCTGGCAGAGG 556
QY      821  CTTGCCAGCCACCAACCGCGCAGC 845
Db      557  GCTGCCAGCGCGCGCGCGCGCACGC 581

```

```

RESULT 15
LOCUS   BF470469
DEFINITION
  UI-M-BH3-avj-b-02-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
ACCESSION
  BF470469
VERSION
  BF470469.1 GI:11539652
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 458)
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
JOURNAL
  Genome Res. 6 (9), 791-806 (1996)
MEDLINE
  9704477
PUBMED
  8889548
COMMENT
  Contact: Chin, H
  National Institute of Mental Health
  6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
  20892-9643, USA
  Tel: 301 443 1706
  Fax: 301 443 9890
  Email: mEst@mail.nih.gov
  cDNA Library Preparation: M. B. Soares Lab Clone distribution.
  Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
  should be noted that Bento Soares is generating a small number of
  additional specialized non-redundant arrays of BMAP cDNAs whose
  availability will be considered under appropriate and limited
  collaborative arrangements
  Seq primer: M13 Reverse.

```

```

FEATURES
  source
  1..458
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="UI-M-BH3-avj-b-02-0-UI"
  /dev_stage="27-32 days"
  /lab_host="DH10B (Life Technologies)"
  /clone_lib="NIH_BMAP_M_S4"
  /note="Vector: pT73D-Pac (Pharmacia) with a modified
  polylinker. Site 1: Not I; Site 2: Eco RI; The
  NIH_BMAP_M_S4 library is a subtracted library of a series,

```

```

BASE COUNT      93 a   138 c   148 g   79 t
ORIGIN
Query Match      18.2%; Score 318; DB 10; Length 458;
Best Local Similarity 97.0%; Pred. No. 1.7e-23;
Matches 324; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY      470  CAGATCCCGTGGAGATGTCCTTTTAAAGGAAGGAAGAGCCGAGCTGGGGCCGCTGA 529
Db      125  CAGATCCCGTGGAGATGTCCTTTTAAAGGAAGGAAGAGCCGAGCTGGGGCCGCTGA 184
QY      530  GGCCTGCGGTACAGAAAGCGCGGACGATCCAGTTGTGGCAGTTTCTCTGGAGCTGC 589
Db      185  GGCCTGCGGTACAGAAAGCGCGGACGATCCAGTTGTGGCAGTTTCTCTGGAGCTGC 244
QY      590  TGGCAGACCGCGGAAACCGCGCTGCATCGCTGGGAGGGCGGCGGAGCTTCAAGC 649
Db      245  TGGCAGATCGCGGAAACCGCGCTTGCATCGCTGGGAGGGCGGCGGAGCTTCAAGC 304
QY      650  TCACCGACCCCGACGAGGTGGCGCGACGCTGGGGGAGCGCGCAAGAGCAAGCCCAATATGA 709
Db      305  TCACCGACCCCGACGAGGTGGCGCGCGCTGGGGGAGCGCGCAAGAGCCCTAAATGA 364
QY      710  ACTAGCAAGCTAAGTCGAGCAGCTGCGCTACTACTACGACAAAACATCATGAGCAAGG 769
Db      365  ACTAGCAAGCTAAGTCGCGGCTGCGCTACTACTACGACAAAACATCATGAGCAAGG 424
QY      770  TGCACGGCAAGCGCTACGCTACCGCTTGGACTT 803
Db      425  TGCACGGCAAGCGCTACGCTACCGCTTGGACTT 458

```

Search completed: November 26, 2003, 16:26:46
Job time : 3866 secs